K102915

510(k) Summary as required by section 807.92(c)

JAN 12 2011

Submitter's name: ViewRay Incorporated
2 Thermo Fisher Way
Oakwood, Ohio 44146
440-703-3210
Fax 440-703-3229

Contact person: Janice Brownlee

Date prepared: September 30, 2010

Trade Name of Device: ViewRay™ Treatment Planning and Delivery System

Common name: Radiation Therapy Treatment Planning and Delivery System

Classification name: Radionuclide Radiation System (21CFR 892.5750, Product Code IWB)

Predicate Device: Varian Medical Systems' Trilogy Mx™ System K092871 and Eclipse™ Treatment Planning System K091492

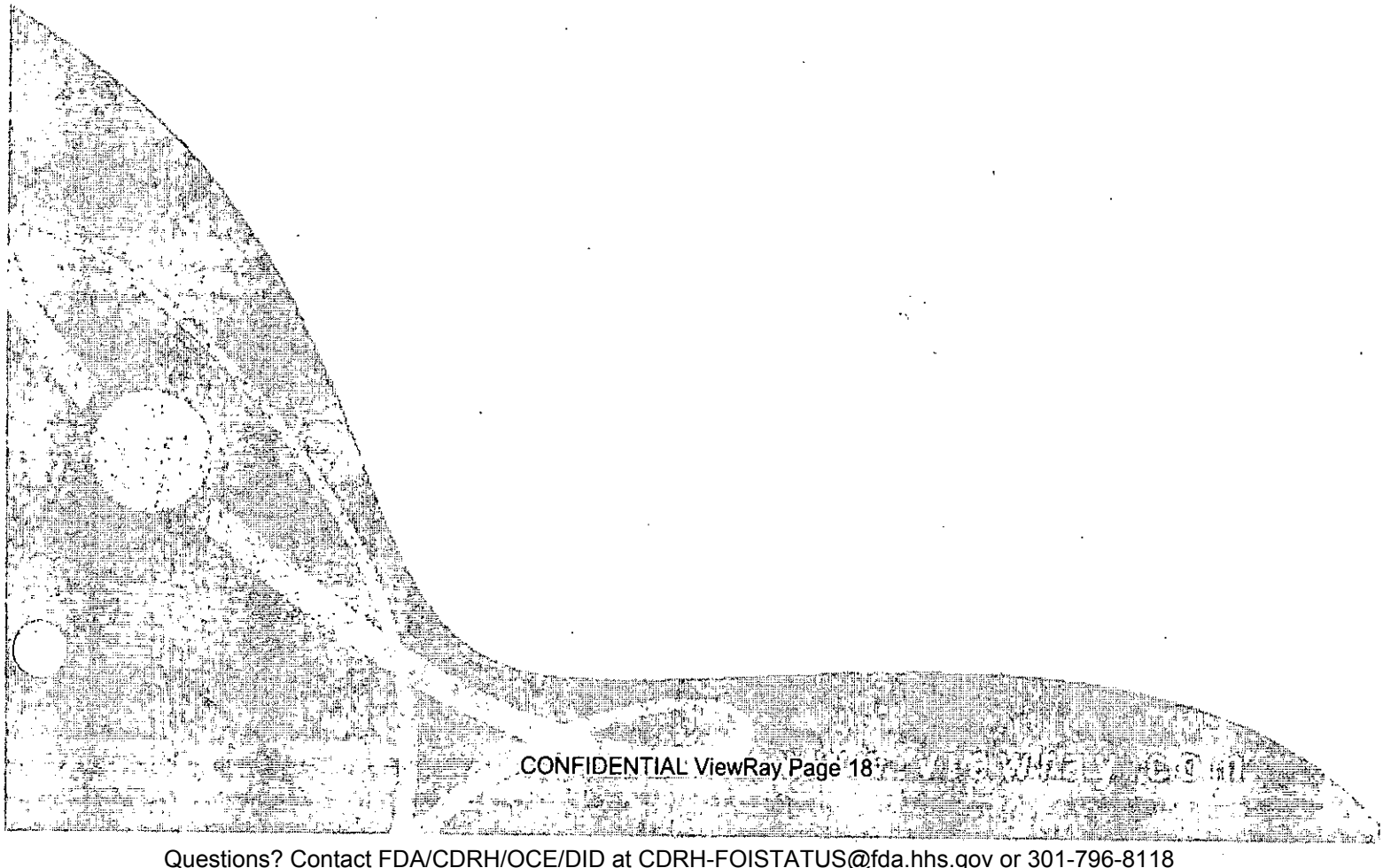
Description: The ViewRay™ Treatment Planning and Delivery System (TPDS) provides tools for planning and delivery of external gamma beam stereotactic radiosurgery and precision radiotherapy for lesions, tumors, and conditions anywhere in the body when radiation treatment is indicated. It is a computer-based device used by trained medical professionals. The Treatment Planning software is only designed to be used on the ViewRay radiation therapy system. The ViewRay TPDS is capable of assisting clinicians in reviewing, prescribing, tracking, and correcting the course of patient treatment using tools for contouring, visualization, data storage, anatomical target monitoring and re-optimization.

The software described in this submission has been designed to conform with applicable sections of IEC 60601-1, IEC 60601-2-11, IEC 62083 and IEC 61217.

Intended use: The ViewRay™ Treatment Planning and Delivery System is intended to be used for planning external beam irradiation with photon beams and delivering stereotactic radiosurgery and precision radiotherapy for lesions, tumors, and conditions anywhere in the body when radiation treatment is indicated, in conjunction with the ViewRay™ System, an MRI image-guided radiation therapy system.

Technological Characteristics compared to predicate device: The ViewRay™Treatment Planning and Delivery System shares many of the technological features and characteristics of the Varian Eclipse planning system, and the treatment delivery features of the Varian Trilogy System. The fundamental technical characteristics are the same as those of the predicate devices and minor differences are described in the comparison chart and discussion provided elsewhere in this 510(k) submission.

Conclusion: The ViewRay™ Treatment Planning and Delivery System has the same intended use, indications for use and user population as the Varian Eclipse™/Trilogy Mx™ System. The ViewRay™ Treatment Planning and Delivery System has most of the features and technological characteristics as the predicate devices, and the few distinguishing characteristics do not raise new types of safety or effectiveness issues.



DEPARTMENT OF HEALTH & HUMAN SERVICES Public Health Service

Food and Drug Administration
10903 New Hampshire Avenue
Document Control Room – WO66-G609
Silver Spring, MD 20993-0002

Ms. Janice Brownlee
VP, Regulatory Affairs and Quality Assurance
ViewRay Incorporated
2 ThermoFisher Way
OAKWOOD VILLAGE OH 44146

JAN 12 2011

Re: K102915
Trade/Device Name: ViewRay™ Treatment Planning and Delivery System
Regulation Number: 21 CFR 892.5750
Regulation Name: Radionuclide radiation therapy system
Regulatory Class: II
Product Code: MUJ
Dated: September 20, 2010
Received: November 1, 2010

Dear Ms. Brownlee:

We have reviewed your Section 510(k) premarket notification of intent to market the device referenced above and have determined the device is substantially equivalent (for the indications for use stated in the enclosure) to legally marketed predicate devices marketed in interstate commerce prior to May 28, 1976, the enactment date of the Medical Device Amendments, or to devices that have been reclassified in accordance with the provisions of the Federal Food, Drug, and Cosmetic Act (Act) that do not require approval of a premarket approval application (PMA). You may, therefore, market the device, subject to the general controls provisions of the Act. The general controls provisions of the Act include requirements for annual registration, listing of devices, good manufacturing practice, labeling, and prohibitions against misbranding and adulteration.

If your device is classified (see above) into class II (Special Controls), it may be subject to such additional controls. Existing major regulations affecting your device can be found in Title 21, Code of Federal Regulations (CFR), Parts 800 to 895. In addition, FDA may publish further announcements concerning your device in the Federal Register.

Please be advised that FDA's issuance of a substantial equivalence determination does not mean that FDA has made a determination that your device complies with other requirements of the Act or any Federal statutes and regulations administered by other Federal agencies. You must comply with all the Act's requirements, including, but not limited to: registration and listing (21 CFR Part 807); labeling (21 CFR Parts 801 and 809); medical device reporting (reporting of

medical device-related adverse events) (21 CFR 803); and good manufacturing practice requirements as set forth in the quality systems (QS) regulation (21 CFR Part 820). This letter will allow you to begin marketing your device as described in your Section 510(k) premarket notification. The FDA finding of substantial equivalence of your device to a legally marketed predicate device results in a classification for your device and thus, permits your device to proceed to the market.

If you desire specific advice for your device on our labeling regulation (21 CFR Parts 801 and 809), please contact the Office of *In Vitro* Diagnostic Device Evaluation and Safety at (301) 796-5450. Also, please note the regulation entitled, "Misbranding by reference to premarket notification" (21 CFR Part 807.97). For questions regarding the reporting of adverse events under the MDR regulation (21 CFR Part 803), please go to http://www.fda.gov/MedicalDevices/Safety/ReportaProblem/default.htm for the CDRH's Office of Surveillance and Biometrics/Division of Postmarket Surveillance.

You may obtain other general information on your responsibilities under the Act from the Division of Small Manufacturers, International and Consumer Assistance at its toll-free number (800) 638-2041 or (301) 796-7100 or at its Internet address http://www.fda.gov/cdrh/industry/support/index.html.

Sincerely yours,

David G. Brown, Ph.D.
Acting Director
Division of Radiological Devices
Office of *In Vitro* Diagnostic Device
Evaluation and Safety
Center for Devices and Radiological Health

Enclosure

1C. Regulatory Sections

Indications for Use

JAN 12 2011

510(k) Number (if known): TBD K102915

Device Name: ViewRay™ Treatment Planning and Delivery System

Use of the ViewRay™ Treatment Planning and Delivery System is indicated for stereotactic radiosurgery and precision radiotherapy for lesions, tumors, and conditions anywhere in the body when radiation treatment is indicated.

Prescription Use X (Part 21 CFR 801 Subpart D) AND/OR Over-The-Counter Use _____ (21 CFR 801 Subpart C)

(PLEASE DO NOT WRITE BELOW THIS LINE-CONTINUE ON ANOTHER PAGE OF NEEDED)

Concurrence of CDRH, Office of Device Evaluation (ODE)

(Division Sign-Off)
Division of Radiological Devices
Office of In Vitro Diagnostic Device Evaluation and Safety

510K K102915

DEPARTMENT OF HEALTH & HUMAN SERVICES Public Health Service

Food and Drug Administration
10903 New Hampshire Avenue
Document Control Room – WO66-G609
Silver Spring, MD 20993-0002

Ms. Janice Brownlee
VP, Regulatory Affairs and Quality Assurance
ViewRay Incorporated
2 ThermoFisher Way
OAKWOOD VILLAGE OH 44146

JAN 12 2011

Re: K102915
Trade/Device Name: ViewRay™ Treatment Planning and Delivery System
Regulation Number: 21 CFR 892.5750
Regulation Name: Radionuclide radiation therapy system
Regulatory Class: II
Product Code: MUJ
Dated: September 20, 2010
Received: November 1, 2010

Dear Ms. Brownlee:

We have reviewed your Section 510(k) premarket notification of intent to market the device referenced above and have determined the device is substantially equivalent (for the indications for use stated in the enclosure) to legally marketed predicate devices marketed in interstate commerce prior to May 28, 1976, the enactment date of the Medical Device Amendments, or to devices that have been reclassified in accordance with the provisions of the Federal Food, Drug, and Cosmetic Act (Act) that do not require approval of a premarket approval application (PMA). You may, therefore, market the device, subject to the general controls provisions of the Act. The general controls provisions of the Act include requirements for annual registration, listing of devices, good manufacturing practice, labeling, and prohibitions against misbranding and adulteration.

If your device is classified (see above) into class II (Special Controls), it may be subject to such additional controls. Existing major regulations affecting your device can be found in Title 21, Code of Federal Regulations (CFR), Parts 800 to 895. In addition, FDA may publish further announcements concerning your device in the Federal Register.

Please be advised that FDA's issuance of a substantial equivalence determination does not mean that FDA has made a determination that your device complies with other requirements of the Act or any Federal statutes and regulations administered by other Federal agencies. You must comply with all the Act's requirements, including, but not limited to: registration and listing (21 CFR Part 807); labeling (21 CFR Parts 801 and 809); medical device reporting (reporting of

medical device-related adverse events) (21 CFR 803); and good manufacturing practice requirements as set forth in the quality systems (QS) regulation (21 CFR Part 820). This letter will allow you to begin marketing your device as described in your Section 510(k) premarket notification. The FDA finding of substantial equivalence of your device to a legally marketed predicate device results in a classification for your device and thus, permits your device to proceed to the market.

If you desire specific advice for your device on our labeling regulation (21 CFR Parts 801 and 809), please contact the Office of *In Vitro* Diagnostic Device Evaluation and Safety at (301) 796-5450. Also, please note the regulation entitled, "Misbranding by reference to premarket notification" (21 CFR Part 807.97). For questions regarding the reporting of adverse events under the MDR regulation (21 CFR Part 803), please go to http://www.fda.gov/MedicalDevices/Safety/ReportaProblem/default.htm for the CDRH's Office of Surveillance and Biometrics/Division of Postmarket Surveillance.

You may obtain other general information on your responsibilities under the Act from the Division of Small Manufacturers, International and Consumer Assistance at its toll-free number (800) 638-2041 or (301) 796-7100 or at its Internet address http://www.fda.gov/cdrh/industry/support/index.html.

Sincerely yours,

David G. Brown, Ph.D.
Acting Director
Division of Radiological Devices
Office of *In Vitro* Diagnostic Device
Evaluation and Safety
Center for Devices and Radiological Health

Enclosure

1C. Regulatory Sections

Indications for Use

JAN 12 [illegible]

510(k) Number (if known): TBD K102915

Device Name: ViewRay™ Treatment Planning and Delivery System

Use of the ViewRay™ Treatment Planning and Delivery System is indicated for stereotactic radiosurgery and precision radiotherapy for lesions, tumors, and conditions anywhere in the body when radiation treatment is indicated.

Prescription Use X AND/OR Over-The-Counter Use _____
(Part 21 CFR 801 Subpart D) (21 CFR 801 Subpart C)

(PLEASE DO NOT WRITE BELOW THIS LINE-CONTINUE ON ANOTHER PAGE OF NEEDED)

Concurrence of CDRH, Office of Device Evaluation (ODE)

(Division Sign-Off)
Division of Radiological Devices
Office of In Vitro Diagnostic Device Evaluation and Safety

510K K102915

DEPARTMENT OF HEALTH & HUMAN SERVICES

Public Health Service

U.S. Food and Drug Administration
Center for Devices and Radiological Health
Document Mail Center ¿ WO66-G609
10903 New Hampshire Avenue
Silver Spring, MD 20993-0002

November 02, 2010

VIEWRAY INCORPORATED
2 THERMOFISHER WAY
OAKWOOD, OHIO 44146
UNITED STATES
ATTN: JANICE BROWNLEE

510k Number: K102915
Received: 11/1/2010
Product: VIEWRAY TREATMENT PLANNING

The Food and Drug Administration (FDA), Center for Devices and Radiological Health (CDRH), has received the Premarket Notification, (510(k)), you submitted in accordance with Section 510(k) of the Federal Food, Drug, and Cosmetic Act(Act) for the above referenced product and for the above referenced 510(k) submitter. Please note, if the 510(k) submitter is incorrect, please notify the 510(k) Staff immediately. We have assigned your submission a unique 510(k) number that is cited above. Please refer prominently to this 510(k) number in all future correspondence that relates to this submission. We will notify you when the processing of your 510(k) has been completed or if any additional information is required. YOU MAY NOT PLACE THIS DEVICE INTO COMMERCIAL DISTRIBUTION UNTIL YOU RECEIVE A LETTER FROM FDA ALLOWING YOU TO DO SO.

Please remember that all correspondence concerning your submission MUST be sent to the Document Mail Center (DMC) at the above letterhead address. Correspondence sent to any address other than the one above will not be considered as part of your official 510(k) submission.

On September 27, 2007, the President signed an act reauthorizing medical device user fees for fiscal years 2008 - 2012. The legislation - the Medical Device User Fee Amendments of 2007 is part of a larger bill, the Food and Drug Amendments Act of 2007. Please visit our website at http://www.fda.gov/MedicalDevices/DeviceRegulationandGuidance/Overview/MedicalDeviceUserFeeandModernizationActMDUFMA/default.htm
for more information regarding fees and FDA review goals. In addition, effective January 2, 2008, any firm that chooses to use a standard in the review of ANY new 510(k) needs to fill out the new standards form (Form 3654) and submit it with their 510(k). The form may be found at http://www.fda.gov/AboutFDA/ReportsManualsForms/Forms/default.htm.

We remind you that Title VIII of the Food and Drug Administration Amendments Act of 2007 (FDAAA) amended the PHS Act by adding new section 402(j) (42 U.S.C. § 282(j)), which expanded the current database known as ClinicalTrials.gov to include mandatory registration and reporting of results for applicable clinical trials of human drugs (including biological products) and devices. Section 402(j) requires that a certification form http://www.fda.gov/AboutFDA/ReportsManualsForms/Forms/default.htm accompany 510(k)/HDE/PMA submissions. The agency has issued a draft guidance titled: "Certifications To Accompany Drug, Biological

Product, and Device Applications/Submissions: Compliance with Section 402(j) of The Public Health Service Act, Added By Title VIII of The Food and Drug Administration Amendments Act of 2007" http://www.fda.gov/MedicalDevices/DeviceRegulationandGuidance/HowtoMarketYourDevice/PremarketSubmissions/PremarketNotification510k/ucm134034.htm. According to the draft guidance, 510(k) submissions that do not contain clinical data do not need the certification form.

Please note the following documents as they relate to 510(k) review: 1) Guidance for Industry and FDA Staff entitled, "Interactive Review for Medical Device Submissions: 510(k)s, Original PMAs, PMA Supplements, Original BLAs and BLA Supplements". This guidance can be found at http://www.fda.gov/MedicalDevices/DeviceRegulationandGuidance/GuidanceDocuments/ucm089402.htm. Please refer to this guidance for information on a formalized interactive review process. 2) Guidance for Industry and FDA Staff entitled, "Format for Traditional and Abbreviated 510(k)s". This guidance can be found at http://www.fda.gov/MedicalDevices/DeviceRegulationandGuidance/GuidanceDocuments/ucm084365.htm. Please refer to this guidance for assistance on how to format an original submission for a Traditional or Abbreviated 510(k).

In all future premarket submissions, we encourage you to provide an electronic copy of your submission. By doing so, you will save FDA resources and may help reviewers navigate through longer documents more easily. Under CDRH's e-Copy Program, you may replace one paper copy of any premarket submission (e.g., 510(k), IDE, PMA, HDE) with an electronic copy. For more information about the program, including the formatting requirements, please visit our web site at http://www.fda.gov/MedicalDevices/DeviceRegulationandGuidance/HowtoMarketYourDevice/PremarketSubmissions/ucm134508.html. In addition, the 510(k) Program Video is now available for viewing on line at http://www.fda.gov/MedicalDevices/DeviceRegulationandGuidance/HowtoMarketYourDevice/PremarketSubmissions/PremarketNotification510k/ucm070201.htm .

Please ensure that whether you submit a 510(k) Summary as per 21 CFR 807.92, or a 510(k) Statement as per 21 CFR 807.93, it meets the content and format regulatory requirements.

Lastly, you should be familiar with the regulatory requirements for medical devices available at Device Advice http://www.fda.gov/MedicalDevices/DeviceRegulationandGuidance/default.htm. If you have questions on the status of your submission, please contact DSMICA at (301)796-7100 or the toll-free number (800)638-2041 , or at their internet address http://www.fda.gov/MedicalDevices/DeviceRegulationandGuidance/default.htm. If you have procedural questions, please contact the 510(k) Staff at (301)796-5640.

Sincerely,

510(k) Staff

DEPARTMENT OF HEALTH & HUMAN SERVICES

Public Health Service

U.S. Food and Drug Administration
Center for Devices and Radiological Health
Document Mail Center ¿ WO66-G609
10903 New Hampshire Avenue
Silver Spring, MD 20993-0002

October 05, 2010

VIEWRAY INCORPORATED
2 THERMOFISHER WAY
OAKWOOD, OHIO 44146
UNITED STATES
ATTN: JANICE BROWNLEE

510k Number: K102915
Received: 10/1/2010
User Fee ID Number:
Product: VIEWRAY TREATMENT PLANNI

The Food and Drug Administration (FDA) Center for Devices and Radiological Health (CDRH), has received the Premarket Notification you submitted in accordance with Section 510(k) of the Federal Food, Drug, and Cosmetic Act (Act) for the above referenced product. We have assigned your submission a unique 510(k) number that is cited above. Please refer prominently to this 510(k) number in all future correspondence that relates to this submission. YOU MAY NOT PLACE THIS DEVICE INTO COMMERCIAL DISTRIBUTION UNTIL YOU RECEIVE A LETTER FROM FDA ALLOWING YOU TO DO SO.

The Federal Food, Drug, and Cosmetic Act (the Act), as amended by the Medical Device User Fee and Modernization Act of 2002 (MDUFMA) and the FDA Amendments Act of 2007 (FDAAA) (Public Law 110-85), authorizes FDA to collect user fees for certain types of 510(k) submissions. The submission cannot be accepted for review until the fee is paid in full ; therefore, the file has been placed on hold. When your user fee payment has been received , review of the 510(k) will resume as of that date. Alternatively, you may request withdrawal of your submission. You now have the option to pay online by credit card. We recommend this form of payment. Credit card payments are directly linked to your user fee cover sheet and are processed the next business day. You may also pay by check. If you choose to mail a check, please send a check to one of the addresses listed below:

By Regular Mail
Food and Drug Administration
P.O. Box 956733
St. Louis, MO 63195-6733.

By Private Courier(e.g.,Fed Ex, UPS, etc.)
U.S. Bank
956733
1005 Convention Plaza
St. Louis, MO 63101

The check should be made out to the Food and Drug Administration referencing the payment identification number, and a copy of the User Fee Cover sheet should be included with the check. A copy of the Medical Device User Fee Cover Sheet should be faxed to CDRH at {ODE_POS_FAX_NUMBER} referencing the 510(k) number if you have not already sent it in with your 510(k) submission. After the FDA has been notified of the receipt of your user fee payment, your 510(k) will be filed and the review will begin. If payment has not been received within 30 days, your 510(k) will be deleted from the system. Additional information on user fees and how to submit your user fee payment may be found at www.fda.gov/cdrh/mdufma/fy09userfee.html. In addition, the 510k Program Video is now available for viewing on line at http://www.fda.gov/MedicalDevices/DeviceRegulationandGuidance/HowtoMarketYourDevice/PremarketSubmissions/PremarketNotification510k/ucm070201.htm.

17

In all future premarket submissions, we encourage you to provide an electronic copy of your submission. By doing so, you will save FDA resources and may help reviewers navigate through longer documents more easily. Under CDRH's e-Copy Program, you may replace one paper copy of any premarket submission (e.g., 510(k), IDE, PMA, or HDE) with an electronic copy. For more information about the program, including the formatting requirements, please visit our web site at http://www.fda.gov/MedicalDevices/DeviceRegulationandGuidance/HowtoMarketYourDevice/PremarketSubmissions/ucm134508.htm.

Please note that since your 510(k) has not been reviewed, additional information may be required during the review process and the file may be placed on hold once again. If you are unsure as to whether or not you need to file a 510k Submission with FDA or what type of submission to submit, you should first telephone the Division of Small Manufacturers, International and Consumer Assistance (DSMICA), for guidance at (301)796-7100 or its toll-fee number (800)638-2041, or contact them at their Internet address http://www.fda.gov/MedicalDevices/DeviceRegulationandGuidance/default.htm, or you may submit a 513(g) request for information regarding classification to the Document Mail Center at the address above. If you have any questions concerning receipt of your payment, please contact Diane Garcia at Diane.Garcia@fda.hhs.gov or directly at (301)796-7200. If you have questions regarding the status of your 510(k) Submission, please contact DSMICA at the numbers or address above.

Sincerely yours,

Diane Garcia
Public Affairs Specialist
Premarket Notification Section
Office of Device Evaluation
Center for Devices and Radiological Health

K102915

1B. ViewRay Introductory Sections

U.S. Food and Drug Administration
Center for Devices and Radiological Health
Document Mail Center - WO66-G609
10903 New Hampshire Avenue
Silver Spring, MD 20993-0002

September 30, 2010

Re: Traditional 510(k) Premarket Notification for ViewRay Treatment Planning and Delivery System

Dear Sir or Madam:

In accordance with Chapter V Section 510(k) Subchapter A of the Federal Food Drugs and Cosmetics Act, and 21 CFR 807 Subpart E, ViewRay Incorporated is submitting this Traditional 510(k).

This is a Traditional submission for the ViewRay Treatment Planning and Delivery System, which meets the classification definition of 21 CFR 892.5750, Radionuclide Radiation Therapy System, Class II. The ViewRay system and its treatment planning software falls under product code IWB, Radiology Radioisotope Therapy Device. This submission is to be reviewed by the CDRH Division of Reproductive, Abdominal and Radiological Devices.

The new, prescription-use ViewRay System device contains software, which is the subject of this submission, and emits radiation. The ViewRay System will consist of a treatment planning system, 0.35T magnetic resonance imager and radiation delivery system. The ViewRay System can use MRI-guided radiation to deliver stereotactic radiosurgery and precision radiotherapy for lesions, tumors, and conditions anywhere in the body when radiation treatment is indicated as prescribed by a physician. The system utilizes the radionuclide Cobalt 60 as the source of radiation, which is a well-established technology.

At the FDA pre-510(k) meeting in October 2009(I070011), the FDA requested that ViewRay submit a 510(k) Premarket Notification for the Treatment Planning and Delivery System software ahead of the MRI-guided radiation therapy system description it is to be used with. A subsequent packet of information covering the integrated hardware and controls for this system will be provided.

ViewRay Incorporated has submitted the required Premarket Notification review fees as required by the FDA. The payment identification number associated with this submission is PIN 50032232, Payment Confirmation Number PCN 10137948. A copy of the User Fee Cover Sheet is included with this submission package.

The information included in this Premarket Notification is detailed in nature. ViewRay considers this information to be confidential and believes the information qualifies for protection under 21 CFR 807.95. Consequently, ViewRay has marked the documents to be protected as "Confidential" and requests the right to review and redact all

This 510(k) Premarket Notification is submitted in duplicate, with two paper copies and an electronic copy as per the FDA's web instructions. The electronic copy is an exact duplicate of the paper copies, except that the documents requiring signature are provided as signed originals accompanying the electronic copy.

Design and Use of the Device

Question	Yes	No
Is the device intended for prescription use (21 CFR 801 Subpart D)?	X	
Is the device intended for over-the-counter use (21 CFR 801 Subpart C)?		X
Does the device contain components derived from a tissue or other biologic source?		X
Is the device provided sterile?		X
Is the device intended for single use?		X
Is the device a reprocessed single use device?		X
If yes, does this device type require reprocessed validation data?		N/A
Does the device contain a drug?		X
Does the device contain a biologic?		X
Does the device use software?	X	
Does the submission include clinical information?		X
Is the device implanted?		X

Please contact me if you have any questions. I can be reached by phone at 440-703-3210 x455, jkbrownlee@viewray.com, or by fax at 440-703-3229.

Respectfully submitted,

Janice Brownlee
VP, Regulatory Affairs and Quality Assurance

VIEWRAY *The Future of Radiotherapy*

Premarket Notification Truthful and Accurate Statement

[As Required by 21 CFR 807.87(k)]

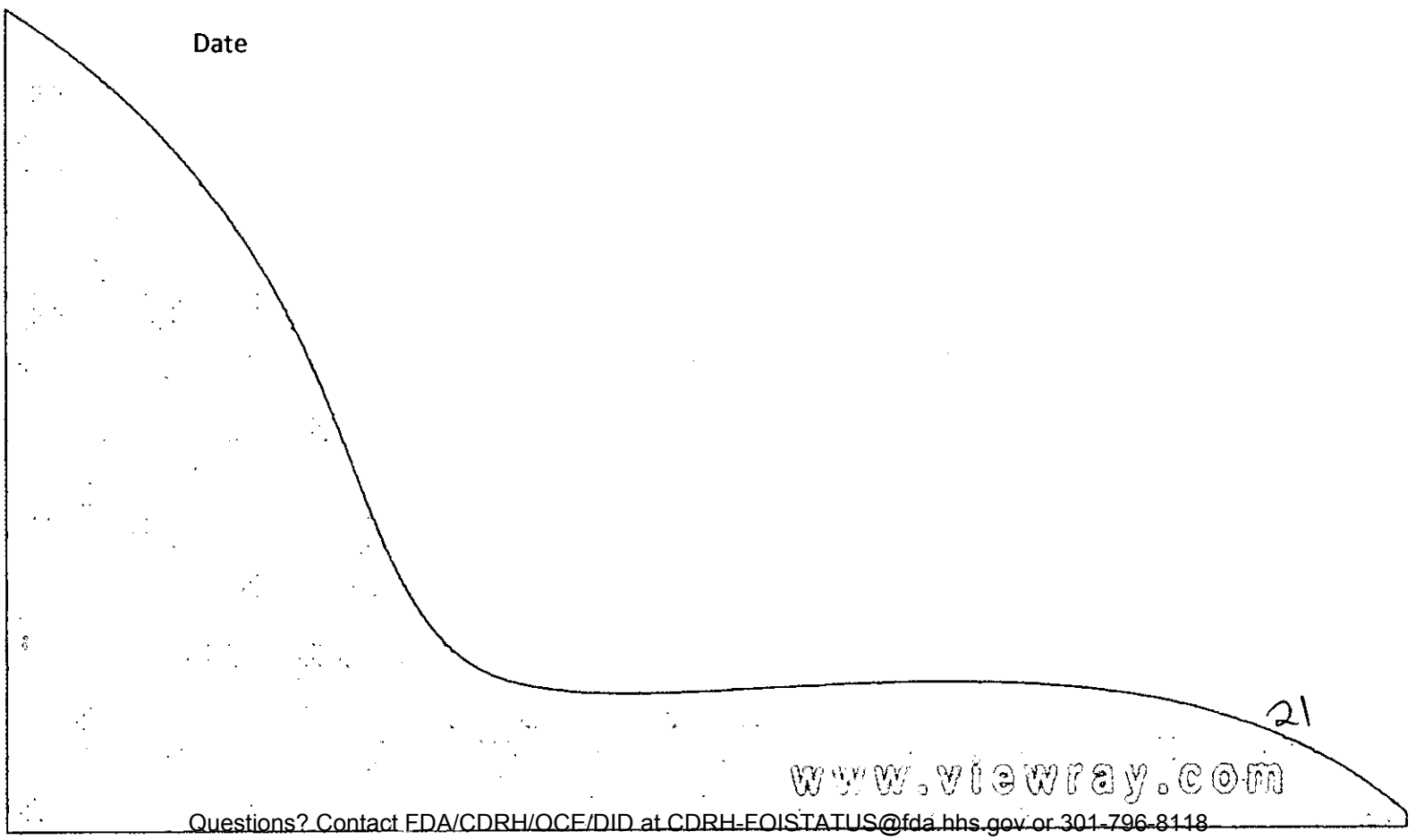
I certify that, in my capacity as VP, Regulatory Affairs and Quality Assurance of ViewRay Incorporated, I believe to the best of my knowledge, that all data and information submitted in the premarket notification are truthful and accurate and that no material fact has been omitted.

Janice Brownlee

Janice Brownlee

9/30/10

Date



See OMB Statement on Reverse. Form Approved: OMB No. 0910-0616, Expiration Date: 10-31-2011

DEPARTMENT OF HEALTH AND HUMAN SERVICES
Food and Drug Administration

Certification of Compliance, under 42 U.S.C. § 282(j)(5)(B), with Requirements of ClinicalTrials.gov Data Bank (42 U.S.C. § 282(j))

(For submission with an application/submission, including amendments, supplements, and resubmissions, under §§ 505, 515, 520(m), or 510(k) of the Federal Food, Drug, and Cosmetic Act or § 351 of the Public Health Service Act.)

SPONSOR / APPLICANT / SUBMITTER INFORMATION

1. NAME OF SPONSOR/APPLICANT/SUBMITTER
ViewRay Incorporated

2. DATE OF THE APPLICATION/SUBMISSION WHICH THIS CERTIFICATION ACCOMPANIES
Sep 30, 2010

3. ADDRESS *(Number, Street, State, and ZIP Code)*
2 ThermoFisher Way, Oakwood, Ohio 44146

4. TELEPHONE AND FAX NUMBERS *(Include Area Code)*
(Tel.) 440-703-3210
(Fax) 440-703-3229

PRODUCT INFORMATION

5. **FOR DRUGS/BIOLOGICS:** Include Any/All Available Established, Proprietary and/or Chemical/Biochemical/Blood/Cellular/Gene Therapy Product Name(s)
FOR DEVICES: Include Any/All Common or Usual Name(s), Classification, Trade or Proprietary or Model Name(s) and/or Model Number(s)
(Attach extra pages as necessary)

Radionuclide Radiation Therapy System, Image-Guided, 21CFR892.5750
ViewRay System, Model 1000

APPLICATION / SUBMISSION INFORMATION

6. TYPE OF APPLICATION/SUBMISSION WHICH THIS CERTIFICATION ACCOMPANIES

☐ IND ☐ NDA ☐ ANDA ☐ BLA ☐ PMA ☐ HDE ☒ 510(k) ☐ PDP ☐ Other

7. INCLUDE IND/NDA/ANDA/BLA/PMA/HDE/510(k)/PDP/OTHER NUMBER *(If number previously assigned)*

8. SERIAL NUMBER ASSIGNED TO APPLICATION/SUBMISSION WHICH THIS CERTIFICATION ACCOMPANIES

CERTIFICATION STATEMENT / INFORMATION

9. CHECK ONLY ONE OF THE FOLLOWING BOXES *(See instructions for additional information and explanation)*

☒ A. I certify that the requirements of 42 U.S.C. § 282(j), Section 402(j) of the Public Health Service Act, enacted by 121 Stat. 823, Public Law 110-85, do not apply because the application/submission which this certification accompanies does not reference any clinical trial.

☐ B. I certify that the requirements of 42 U.S.C. § 282(j), Section 402(j) of the Public Health Service Act, enacted by 121 Stat. 823, Public Law 110-85, do not apply to any clinical trial referenced in the application/submission which this certification accompanies.

☐ C. I certify that the requirements of 42 U.S.C. § 282(j), Section 402(j) of the Public Health Service Act, enacted by 121 Stat. 823, Public Law 110-85, apply to one or more of the clinical trials referenced in the application/submission which this certification accompanies and that those requirements have been met.

10. IF YOU CHECKED BOX C, IN NUMBER 9, PROVIDE THE NATIONAL CLINICAL TRIAL (NCT) NUMBER(S) FOR ANY "APPLICABLE CLINICAL TRIAL(S)," UNDER 42 U.S.C. § 282(j)(1)(A)(i), SECTION 402(j)(1)(A)(i) OF THE PUBLIC HEALTH SERVICE ACT, REFERENCED IN THE APPLICATION/ SUBMISSION WHICH THIS CERTIFICATION ACCOMPANIES *(Attach extra pages as necessary)*

NCT Number(s):

The undersigned declares, to the best of her/his knowledge, that this is an accurate, true, and complete submission of information. I understand that the failure to submit the certification required by 42 U.S.C. § 282(j)(5)(B), section 402(j)(5)(B) of the Public Health Service Act, and the knowing submission of a false certification under such section are prohibited acts under 21 U.S.C. § 331, section 301 of the Federal Food, Drug, and Cosmetic Act.
Warning: A willfully and knowingly false statement is a criminal offense, U.S. Code, title 18, section 1001.

11. SIGNATURE OF SPONSOR/APPLICANT/SUBMITTER OR AN AUTHORIZED REPRESENTATIVE (Sign)
Janice Brownlee

12. NAME AND TITLE OF THE PERSON WHO SIGNED IN NO. 11
(Name) Janice Brownlee
(Title) VP, Regulatory Affairs and Quality Assurance

13. ADDRESS *(Number, Street, State, and ZIP Code) (of person identified in Nos. 11 and 12)*
ViewRay Incorporated
2 ThermoFisher Way
Oakwood, OH 44146

14. TELEPHONE AND FAX NUMBERS *(Include Area Code)*
(Tel.) 440-703-3210
(Fax) 440-703-3229

15. DATE OF CERTIFICATION
9/30/10

PSC Graphics: (301) 443-1090 EF

ViewRay, Inc.

ViewRay Treatment Planning
and Delivery System

COPY 2
1 of 1

Traditional 510(k) Submission
30-SEP-2010

25

Traditional 510(k) Premarket Notification for VIEWRAY™ TREATMENT PLANNING and DELIVERY SYSTEM

K-47

Section 1

Introduction-
Regulatory Forms

1A. Cover Sheets

MDUFMA User Fee Cover Sheet:

Form Approved: OMB No. 0910-0625 Expiration Date: March 31, 2012. See Instructions for OMB Statement.

DEPARTMENT OF HEALTH AND HUMAN SERVICES FOOD AND DRUG ADMINISTRATION **DEVICE FACILITY USER FEE**	PAYMENT IDENTIFICATION NUMBER: (b)(4) Include the Payment Identification Number (PIN) with payment.

The following actions must be taken to properly submit your payment:

1. To submit payment, please select one of the following options:

A. To pay electronically using ACH (electronic check from a US bank) or a credit card, please select the "Pay Now" option.

B. To pay using a check drawn on a US bank in US dollars, please follow these instructions:

- Make check payable to the Food and Drug Administration
- Write the payment identification number (PIN) on the check
- Mail check and a printed copy of the order to:

  Food and Drug Administration
  P.O. Box 70961
  Charlotte, NC 28272-0961

OR

- For checks sent by courier, mail the check and printed copy of the order to:

  Wells Fargo Bank
  Attn: Food and Drug Administration Lockbox 70961
  1525 West WT Harris Blvd., Room D1113-022
  Charlotte, NC 28262

Note: This Wells Fargo Bank address is for courier delivery only; do not send mail to this address.

C. To pay by wire transfer, please read the following:

You are responsible to pay any administrative costs associated with the processing of a wire transfer. Contact your bank or financial institution regarding the additional fees.

US Department of Treasury
TREAS NYC
33 Liberty Street
New York, NY 10045

FDA Deposit Account Number: 75060099
Beneficiary: Food and Drug Administration, 1350 Piccard Drive, Suite 200A, Rockville, MD 20850
US Department of Treasury routing/transit number: 021030004
SWIFT Number: FRNYUS33

You must include the user fee payment identification number (PIN), and ensure that the fee that your bank will charge for the wire transfer is added to your fee payment.

2. Company Name and Address	3. Contact Name
ViewRay Incorporated 2 Thermo Fisher Way Oakwood OH 44146 US	Janice Brownlee
2.1 Employer Identification Number (EIN) *****8429	3.1 E-mail Address jkbrownlee@viewray.com
	3.2 Telephone Number 440-703-3210
	3.3 Fax Number

4. PIN-PCN (*Payment Identification Number-Payment Confirmation Number*):

(b)(4)

5. Amount Due:

(b)(4) 21-Sep-2010

Close Print Order

Form FDA 3601 (01/2007)

UFMA Small Business Qualification Certification:

DEPARTMENT OF HEALTH & HUMAN SERVICES

Food and Drug Administration
10903 New Hampshire Avenue
Room 4608, Building 66
Silver Spring, MD 20993-0002

September 17, 2010

ViewRay Incorporated
2 Thermo Fisher Way
Oakwood, OH 44146

To: Janice Brownlee, VP, Regulatory Affairs and Quality Assurance
Subject: Request for Certification Qualification as a Small Business

Small Business Decision Number: SBD100256
Expires: September 30, 2010

This responds to your request for eligibility for Small Business Qualification on Form FDA 3602.

After review of your submission, I am pleased to inform you that your firm does qualify under MDUFMA as a small business for reduced or waived fees for medical device submissions made during the fiscal year 2010

Please include your Small Business Decision Number (see above) whenever you submit a Medical Device User Fee Coversheet (form FDA 3601). This form is available at http://www.fda.gov/oc/mdufma/coversheet.html. In completing the form, the Business EIN Number and the user fee organization number 249593 must correspond to the Business name in the address above.

If you are registering as a new user to the User Fee System, please use this organization number to register as an existing organization. If you currently have a User Fee account and the organization number in your profile does not match this organization number, please contact the User Fees Help Desk for further assistance at 301-796-7200 or userfeesupport@fda.hhs.gov.

Your Small Business status expires at the close of business September 30, 2010. FDA will provide information on how to qualify as a Small Business for FY 2011 in a Federal Register Notice to be published on or around August 1, 2010. We will also provide this information on our MDUFMA website, at: http://www.fda.gov/cdrh/mdufma/.

Connie Daly
Small Business Decision (SBD) Reviewer
Division of Small Manufacturers, International and
Consumer Assistance
Office of Communication, Education, and Radiation Programs
Center for Devices and Radiological Health

CDRH Premarket Review Submission Cover Sheet (5 pages):

DEPARTMENT OF HEALTH AND HUMAN SERVICES
FOOD AND DRUG ADMINISTRATION

CDRH PREMARKET REVIEW SUBMISSION COVER SHEET

Form Approval
OMB No. 0910-0120
Expiration Date: August 31, 2010.
See OMB Statement on page 5.

Date of Submission	User Fee Payment ID Number	FDA Submission Document Number *(if known)*
09/30/2010	(b)(4)	TBD

SECTION A — TYPE OF SUBMISSION

PMA	PMA & HDE Supplement	PDP	510(k)	Meeting
☐ Original Submission	☐ Regular (180 day)	☐ Original PDP	☒ Original Submission:	☐ Pre-510(K) Meeting
☐ Premarket Report	☐ Special	☐ Notice of Completion	☒ Traditional	☐ Pre-IDE Meeting
☐ Modular Submission	☐ Panel Track (PMA Only)	☐ Amendment to PDP	☐ Special	☐ Pre-PMA Meeting
☐ Amendment	☐ 30-day Supplement		☐ Abbreviated (Complete section I, Page 5)	☐ Pre-PDP Meeting
☐ Report	☐ 30-day Notice		☐ Additional Information	☐ Day 100 Meeting
☐ Report Amendment	☐ 135-day Supplement		☐ Third Party	☐ Agreement Meeting
☐ Licensing Agreement	☐ Real-time Review			☐ Determination Meeting
	☐ Amendment to PMA & HDE Supplement			☐ Other *(specify)*:
	☐ Other			

IDE	Humanitarian Device Exemption (HDE)	Class II Exemption Petition	Evaluation of Automatic Class III Designation (De Novo)	Other Submission
☐ Original Submission	☐ Original Submission	☐ Original Submission	☐ Original Submission	☐ 513(g)
☐ Amendment	☐ Amendment	☐ Additional Information	☐ Additional Information	☐ Other *(describe submission)*:
☐ Supplement	☐ Supplement			
	☐ Report			
	☐ Report Amendment			

Have you used or cited Standards in your submission? ☒ Yes ☐ No *(If Yes, please complete Section I, Page 5)*

SECTION B — SUBMITTER, APPLICANT OR SPONSOR

Company / Institution Name	Establishment Registration Number *(if known)*		
ViewRay Incorporated	TBD		
Division Name *(if applicable)*	**Phone Number *(including area code)***		
	440-703-3210		
Street Address	**FAX Number *(including area code)***		
2 ThermoFisher Way	440-703-3229		
City	**State / Province**	**ZIP/Postal Code**	**Country**
Oakwood Village	Ohio	44146	USA
Contact Name			
Janice Brownlee			
Contact Title	**Contact E-mail Address**		
VP, Regulatory Affairs and Quality Assurance	jkbrownlee@viewray.com		

SECTION C — APPLICATION CORRESPONDENT (e.g., consultant, if different from above)

Company / Institution Name			
Division Name *(if applicable)*	**Phone Number *(including area code)***		
Street Address	**FAX Number *(including area code)***		
City	**State / Province**	**ZIP Code**	**Country**
Contact Name			
Contact Title	**Contact E-mail Address**		

FORM FDA 3514 (3/08) Page 1 of 5 Pages

SECTION D1 REASON FOR APPLICATION - PMA, PDP, OR HDE

- ☐ New Device
- ☐ Withdrawal
- ☐ Additional or Expanded Indications
- ☐ Request for Extension
- ☐ Post-approval Study Protocol
- ☐ Request for Applicant Hold
- ☐ Request for Removal of Applicant Hold
- ☐ Request to Remove or Add Manufacturing Site

- ☐ Change in design, component, or specification:
  - ☐ Software / Hardware
  - ☐ Color Additive
  - ☐ Material
  - ☐ Specifications
  - ☐ Other *(specify below)*

- ☐ Location change:
  - ☐ Manufacturer
  - ☐ Sterilizer
  - ☐ Packager

- ☐ Process change:
  - ☐ Manufacturing ☐ Packaging
  - ☐ Sterilization
  - ☐ Other *(specify below)*

- ☐ Labeling change:
  - ☐ Indications
  - ☐ Instructions
  - ☐ Performance Characteristics
  - ☐ Shelf Life
  - ☐ Trade Name
  - ☐ Other *(specify below)*

- ☐ Report Submission:
  - ☐ Annual or Periodic
  - ☐ Post-approval Study
  - ☐ Adverse Reaction
  - ☐ Device Defect
  - ☐ Amendment

- ☐ Response to FDA correspondence:

- ☐ Change in Ownership
- ☐ Change in Correspondent
- ☐ Change of Applicant Address

- ☐ Other Reason *(specify)*:

SECTION D2 REASON FOR APPLICATION - IDE

- ☐ New Device
- ☐ New Indication
- ☐ Addition of Institution
- ☐ Expansion / Extension of Study
- ☐ IRB Certification
- ☐ Termination of Study
- ☐ Withdrawal of Application
- ☐ Unanticipated Adverse Effect
- ☐ Notification of Emergency Use
- ☐ Compassionate Use Request
- ☐ Treatment IDE
- ☐ Continued Access

- ☐ Change in:
  - ☐ Correspondent / Applicant
  - ☐ Design / Device
  - ☐ Informed Consent
  - ☐ Manufacturer
  - ☐ Manufacturing Process
  - ☐ Protocol - Feasibility
  - ☐ Protocol - Other
  - ☐ Sponsor

- ☐ Report submission:
  - ☐ Current Investigator
  - ☐ Annual Progress Report
  - ☐ Site Waiver Report
  - ☐ Final

- ☐ Response to FDA Letter Concerning:
  - ☐ Conditional Approval
  - ☐ Deemed Approved
  - ☐ Deficient Final Report
  - ☐ Deficient Progress Report
  - ☐ Deficient Investigator Report
  - ☐ Disapproval
  - ☐ Request Extension of Time to Respond to FDA
  - ☐ Request Meeting
  - ☐ Request Hearing

- ☐ Other Reason *(specify)*:

SECTION D3 REASON FOR SUBMISSION - 510(k)

- ☒ New Device
- ☐ Additional or Expanded Indications
- ☐ Change in Technology

- ☐ Other Reason *(specify)*:

FORM FDA 3514 (3/08) Page 2 of 5 Pages

SECTION E ADDITIONAL INFORMATION ON 510(K) SUBMISSIONS

Product codes of devices to which substantial equivalence is claimed

1	MUJ	2	IYE	3		4	
5		6		7		8	

Summary of, or statement concerning, safety and effectiveness information

☒ 510 (k) summary attached
☐ 510 (k) statement

Information on devices to which substantial equivalence is claimed *(if known)*

	510(k) Number		Trade or Proprietary or Model Name		Manufacturer
1	K091492, FDA clearance date 6/18/09	1	Varian Eclipse™	1	Varian Medical Systems
2	K092871, FDA clearance date 11/30/09	2	Varian Trilogy™	2	Varian Medical Systems
3		3		3	
4		4		4	
5		5		5	
6		6		6	

SECTION F PRODUCT INFORMATION - APPLICATION TO ALL APPLICATIONS

Common or usual name or classification name

Image-Guided Radionuclide Radiation Therapy System

	Trade or Proprietary or Model Name for This Device		Model Number
1	ViewRay Treatment Planning and Delivery System	1	
2		2	
3		3	
4		4	
5		5	

FDA document numbers of all prior related submissions *(regardless of outcome)*

1 I070011	2 I070011/S001-006	3	4	5	6
7	8	9	10	11	12

Data Included in Submission

☒ Laboratory Testing ☐ Animal Trials ☐ Human Trials

SECTION G PRODUCT CLASSIFICATION - APPLICATION TO ALL APPLICATIONS

Product Code	C.F.R. Section *(if applicable)*	Device Class
IWB	21CFR 892.5750 Radionuclide Radiation Therapy System	☐ Class I ☒ Class II
Classification Panel		☐ Class III ☐ Unclassified
Division of Reproductive, Abdominal and Radiological Devices		

Indications *(from labeling)*

Use of the ViewRay™ Treatment Planning and Delivery System is indicated for stereotactic radiosurgery and precision radiotherapy for lesions, tumors, and conditions anywhere in the body when radiation treatment is indicated.

Note: Submission of this information does not affect the need to submit a 2891 or 2891a Device Establishment Registration form.	FDA Document Number *(if known)*

SECTION H — MANUFACTURING / PACKAGING / STERILIZATION SITES RELATING TO A SUBMISSION

☒ Original ☐ Add ☐ Delete	Facility Establishment Identifier (FEI) Number: Federal ID # 20-0818429	☒ Manufacturer ☐ Contract Manufacturer	☐ Contract Sterilizer ☐ Repackager / Relabeler
Company / Institution Name: ViewRay Incorporated		Establishment Registration Number: TBD	
Division Name *(if applicable)*		Phone Number *(including area code)*: 440-703-3210	
Street Address: 2 ThermoFisher Way		FAX Number *(including area code)*	
City: Oakwood	State / Province: OH	ZIP Code: 44146	Country: USA
Contact Name: Janice Brownlee	Contact Title: VP, Regulatory Affairs/Quality Assurance	Contact E-mail Address: jkbrownlee@viewray.com	

☐ Original ☐ Add ☐ Delete	Facility Establishment Identifier (FEI) Number	☐ Manufacturer ☐ Contract Manufacturer	☐ Contract Sterilizer ☐ Repackager / Relabeler
Company / Institution Name		Establishment Registration Number	
Division Name *(if applicable)*		Phone Number *(including area code)*	
Street Address		FAX Number *(including area code)*	
City	State / Province	ZIP Code	Country
Contact Name	Contact Title	Contact E-mail Address	

☐ Original ☐ Add ☐ Delete	Facility Establishment Identifier (FEI) Number	☐ Manufacturer ☐ Contract Manufacturer	☐ Contract Sterilizer ☐ Repackager / Relabeler
Company / Institution Name		Establishment Registration Number	
Division Name *(if applicable)*		Phone Number *(including area code)*	
Street Address		FAX Number *(including area code)*	
City	State / Province	ZIP Code	Country
Contact Name	Contact Title	Contact E-mail Address	

SECTION I	UTILIZATION OF STANDARDS				
Note: Complete this section if your application or submission cites standards or includes a *"Declaration of Conformity to a Recognized Standard"* statement.					
	Standards No.	Standards Organization	Standards Title	Version	Date
1	Recognition Number 12-190: 61217	IEC	Radiotherapy equipment- coordinates,movements and scales	1.2	01/01/2008
	Standards No.	Standards Organization	Standards Title	Version	Date
2	Recognition number 12-122: 62083	IEC	Medical electrical equipment:Requirements for the safety of radiotherapy treatment planning systems	1.0	01/01/2000
	Standards No.	Standards Organization	Standards Title	Version	Date
3	Recognition Number 5-4: 60601-1	IEC	Medical electrical equipment-Part1:General Requirements for Safety,1988; Amendment 1,1991-11, Amendment 2. 1995	1988,1991,1995	01/01/1995
	Standards No.	Standards Organization	Standards Title	Version	Date
4	Recognition Number:12-133: 60601-2-11	IEC	Particular Requirements for the safety of gamma beam therapy equipment	1997 and 2004	01/01/1997
	Standards No.	Standards Organization	Standards Title	Version	Date
5					
	Standards No.	Standards Organization	Standards Title	Version	Date
6					
	Standards No.	Standards Organization	Standards Title	Version	Date
7					
Please include any additional standards to be cited on a separate page.					

Public reporting burden for this collection of information is estimated to average 0.5 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing reviewing the collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden to:

Department of Health and Human Services
Food and Drug Administration
Office of the Chief Information Officer (HFA-710)
5600 Fishers Lane
Rockville, Maryland 20857

An agency may not conduct or sponsor, and a person is not required to respond to, a collection of information unless it displays a currently valid OMB control number.

FORM FDA 3514 (3/08) Page 5 of 5 Pages

1B. ViewRay™ Introductory Sections

U.S. Food and Drug Administration
Center for Devices and Radiological Health
Document Mail Center - WO66-G609
10903 New Hampshire Avenue
Silver Spring, MD 20993-0002

September 30, 2010

Re: Traditional 510(k) Premarket Notification for ViewRay™ Treatment Planning and Delivery System

Dear Sir or Madam:

In accordance with Chapter V Section 510(k) Subchapter A of the Federal Food Drugs and Cosmetics Act, and 21 CFR 807 Subpart E, ViewRay™ Incorporated is submitting this Traditional 510(k).

This is a Traditional submission for the ViewRay™ Treatment Planning and Delivery System, which meets the classification definition of 21 CFR 892.5750, Radionuclide Radiation Therapy System, Class II. The ViewRay™ system and its treatment planning software falls under product code IWB, Radiology Radioisotope Therapy Device. This submission is to be reviewed by the CDRH Division of Reproductive, Abdominal and Radiological Devices.

The new, prescription-use ViewRay™ System device contains software, which is the subject of this submission, and emits radiation. The ViewRay™ System will consist of a treatment planning system, 0.35T magnetic resonance imager and radiation delivery system. The ViewRay™ System can use MRI-guided radiation to deliver stereotactic radiosurgery and precision radiotherapy for lesions, tumors, and conditions anywhere in the body when radiation treatment is indicated as prescribed by a physician. The system utilizes the radionuclide Cobalt 60 as the source of radiation, which is a well-established technology.

At the FDA pre-510(k) meeting in October 2009(I070011), the FDA requested that ViewRay™ submit a 510(k) Premarket Notification for the Treatment Planning and Delivery System software ahead of the MRI-guided radiation therapy system description it is to be used with. A subsequent packet of information covering the integrated hardware and controls for this system will be provided.

ViewRay™ Incorporated has submitted the required Premarket Notification review fees as required by the FDA. The payment identification number associated with this submission is PIN 5(b)(4), Payment Confirmation Number PCN(b)(4). A copy of the User Fee Cover Sheet is included with this submission package.

The information included in this Premarket Notification is detailed in nature. ViewRay™ considers this information to be confidential and believes the information qualifies for protection under 21 CFR 807.95. Consequently, ViewRay™ has marked the documents to be protected as "Confidential" and requests the right to review and redact all

This 510(k) Premarket Notification is submitted in duplicate, with two paper copies and an electronic copy as per the FDA's web instructions. The electronic copy is an exact duplicate of the paper copies, except that the documents requiring signature are provided as signed originals accompanying the electronic copy.

Design and Use of the Device

Question	Yes	No
Is the device intended for prescription use (21 CFR 801 Subpart D)?	X	
Is the device intended for over-the-counter use (21 CFR 801 Subpart C)?		X
Does the device contain components derived from a tissue or other biologic source?		X
Is the device provided sterile?		X
Is the device intended for single use?		X
Is the device a reprocessed single use device?		X
If yes, does this device type require reprocessed validation data?		N/A
Does the device contain a drug?		X
Does the device contain a biologic?		X
Does the device use software?	X	
Does the submission include clinical information?		X
Is the device implanted?		X

Please contact me if you have any questions. I can be reached by phone at 440-703-3210 x455, jkbrownlee@viewray.com, or by fax at 440-703-3229.

Respectfully submitted,

Janice Brownlee
VP, Regulatory Affairs and Quality Assurance

VIEWRAY INCORPORATED *The Future of Radiotherapy*

VIEWRAY™ TREATMENT PLANNING AND DELIVERY SYSTEM SOFTWARE DISCUSSION

TABLE OF CONTENTS

37

CDRH Screening Checklist for Traditional Premarket Notification:

Requirement	Section
MDUFMA Cover Sheet	Section 1A
CDRH Premarket Review Submission Cover Sheet	Section 1A
510(k) Cover Letter	Section 1B or accompanying electronic copy
Indications for Use Statement	Section 1C
510(k) Summary or 510(k) Statement	Section 1C
Truthful and Accuracy Statement	Section 1C or accompanying electronic copy
Certificate of Compliance with Requirements of Clinical Trials Data Bank	Section 1C or accompanying electronic copy
Financial Certification or Disclosure Statement	N/A
Declarations of Conformity and Summary Reports-FORM FDA 3654, *Standards Data Report for 510(k)s*	Section 1C
Executive Summary	Section 1C
Device Description	Section 2
Substantial Equivalence Discussion	Section 3
Proposed Labeling	Section 5 and Attachments 5A
Sterilization/Shelf Life	N/A
Biocompatibility	N/A
Software	Entire document
Electromagnetic Compatibility/Electrical Safety	N/A
Performance Testing – Bench	Section 6 and Attachments 6A, 6B
Performance Testing – Animal	N/A
Performance Testing – Clinical	N/A

1C. Regulatory Sections

Indications for Use

510(k) Number (if known): TBD K102915

Device Name: ViewRay™ Treatment Planning and Delivery System

Use of the ViewRay™ Treatment Planning and Delivery System is indicated for stereotactic radiosurgery and precision radiotherapy for lesions, tumors, and conditions anywhere in the body when radiation treatment is indicated.

Prescription Use X (Part 21 CFR 801 Subpart D) AND/OR Over-The-Counter Use _____ (21 CFR 801 Subpart C)

(PLEASE DO NOT WRITE BELOW THIS LINE-CONTINUE ON ANOTHER PAGE OF NEEDED)

Concurrence of CDRH, Office of Device Evaluation (ODE)

510(k) Summary as required by section 807.92(c)

Submitter's name: ViewRay Incorporated
2 Thermo Fisher Way
Oakwood, Ohio 44146
440-703-3210
Fax 440-703-3229

Contact person: Janice Brownlee

Date prepared: September 30, 2010

Trade Name of Device: ViewRay™ Treatment Planning and Delivery System

Common name: Radiation Therapy Treatment Planning and Delivery System

Classification name: Radionuclide Radiation System (21CFR 892.5750, Product Code IWB)

Predicate Device: Varian Medical Systems' Trilogy Mx™ System K092871 and Eclipse™ Treatment Planning System K091492

Description: The ViewRay™ Treatment Planning and Delivery System (TPDS) provides tools for planning and delivery of external gamma beam stereotactic radiosurgery and precision radiotherapy for lesions, tumors, and conditions anywhere in the body when radiation treatment is indicated. It is a computer-based device used by trained medical professionals. The Treatment Planning software is only designed to be used on the ViewRay radiation therapy system. The ViewRay TPDS is capable of assisting clinicians in reviewing, prescribing, tracking, and correcting the course of patient treatment using tools for contouring, visualization, data storage, anatomical target monitoring and re-optimization.

The software described in this submission has been designed to conform with applicable sections of IEC 60601-1, IEC 60601-2-11, IEC 62083 and IEC 61217.

Intended use: The ViewRay™ Treatment Planning and Delivery System is intended to be used for planning external beam irradiation with photon beams and delivering stereotactic radiosurgery and precision radiotherapy for lesions, tumors, and conditions anywhere in the body when radiation treatment is indicated, in conjunction with the ViewRay™ System, an MRI image-guided radiation therapy system.

Technological Characteristics compared to predicate device: The ViewRay™Treatment Planning and Delivery System shares many of the technological features and characteristics of the Varian Eclipse planning system, and the treatment delivery features of the Varian Trilogy System. The fundamental technical characteristics are the same as those of the predicate devices and minor differences are described in the comparison chart and discussion provided elsewhere in this 510(k) submission.

Conclusion: The ViewRay™ Treatment Planning and Delivery System has the same intended use, indications for use and user population as the Varian Eclipse™/Trilogy Mx™ System. The ViewRay™ Treatment Planning and Delivery System has most of the features and technological characteristics as the predicate devices, and the few distinguishing characteristics do not raise new types of safety or effectiveness issues.

Premarket Notification Truthful and Accurate Statement

[As Required by 21 CFR 807.87(k)]

I certify that, in my capacity as VP, Regulatory Affairs and Quality Assurance of ViewRay Incorporated, I believe to the best of my knowledge, that all data and information submitted in the premarket notification are truthful and accurate and that no material fact has been omitted.

Janice Brownlee

Janice Brownlee

Sept. 30, 2010

Date

Certificate of Compliance FDA Form 3674(1 page)

See OMB Statement on Reverse. Form Approved. OMB No. 0910-0616. Expiration Date: 10-31-2011

DEPARTMENT OF HEALTH AND HUMAN SERVICES
Food and Drug Administration

Certification of Compliance, under 42 U.S.C. § 282(j)(5)(B), with Requirements of ClinicalTrials.gov Data Bank (42 U.S.C. § 282(j))

(For submission with an application/submission, including amendments, supplements, and resubmissions, under §§ 505, 515, 520(m), or 510(k) of the Federal Food, Drug, and Cosmetic Act or § 351 of the Public Health Service Act.)

SPONSOR / APPLICANT / SUBMITTER INFORMATION

1. NAME OF SPONSOR/APPLICANT/SUBMITTER
ViewRay Incorporated

2. DATE OF THE APPLICATION/SUBMISSION WHICH THIS CERTIFICATION ACCOMPANIES
Sep 30, 2010

3. ADDRESS *(Number, Street, State, and ZIP Code)*
2 ThermoFisher Way, Oakwood, Ohio 44146

4. TELEPHONE AND FAX NUMBERS *(Include Area Code)*
(Tel.) 440-703-3210
(Fax) 440-703-3229

PRODUCT INFORMATION

5. FOR DRUGS/BIOLOGICS: Include Any/All Available Established, Proprietary and/or Chemical/Biochemical/Blood/Cellular/Gene Therapy Product Name(s)
FOR DEVICES: Include Any/All Common or Usual Name(s), Classification, Trade or Proprietary or Model Name(s) and/or Model Number(s)
(Attach extra pages as necessary)
Radionuclide Radiation Therapy System, Image-Guided, 21CFR892.5750
ViewRay System, Model 1000

APPLICATION / SUBMISSION INFORMATION

6. TYPE OF APPLICATION/SUBMISSION WHICH THIS CERTIFICATION ACCOMPANIES
☐ IND ☐ NDA ☐ ANDA ☐ BLA ☐ PMA ☐ HDE ☒ 510(k) ☐ PDP ☐ Other

7. INCLUDE IND/NDA/ANDA/BLA/PMA/HDE/510(k)/PDP/OTHER NUMBER *(If number previously assigned)*

8. SERIAL NUMBER ASSIGNED TO APPLICATION/SUBMISSION WHICH THIS CERTIFICATION ACCOMPANIES

CERTIFICATION STATEMENT / INFORMATION

9. CHECK ONLY ONE OF THE FOLLOWING BOXES *(See instructions for additional information and explanation)*

☒ A. I certify that the requirements of 42 U.S.C. § 282(j), Section 402(j) of the Public Health Service Act, enacted by 121 Stat. 823, Public Law 110-85, do not apply because the application/submission which this certification accompanies does not reference any clinical trial.

☐ B. I certify that the requirements of 42 U.S.C. § 282(j), Section 402(j) of the Public Health Service Act, enacted by 121 Stat. 823, Public Law 110-85, do not apply to any clinical trial referenced in the application/submission which this certification accompanies.

☐ C. I certify that the requirements of 42 U.S.C. § 282(j), Section 402(j) of the Public Health Service Act, enacted by 121 Stat. 823, Public Law 110-85, apply to one or more of the clinical trials referenced in the application/submission which this certification accompanies and that those requirements have been met.

10. IF YOU CHECKED BOX C. IN NUMBER 9, PROVIDE THE NATIONAL CLINICAL TRIAL (NCT) NUMBER(S) FOR ANY "APPLICABLE CLINICAL TRIAL(S)." UNDER 42 U.S.C. § 282(j)(1)(A)(i). SECTION 402(j)(1)(A)(i) OF THE PUBLIC HEALTH SERVICE ACT, REFERENCED IN THE APPLICATION/ SUBMISSION WHICH THIS CERTIFICATION ACCOMPANIES *(Attach extra pages as necessary)*

NCT Number(s):

The undersigned declares, to the best of her/his knowledge, that this is an accurate, true, and complete submission of information. I understand that the failure to submit the certification required by 42 U.S.C. § 282(j)(5)(B), section 402(j)(5)(B) of the Public Health Service Act, and the knowing submission of a false certification under such section are prohibited acts under 21 U.S.C. § 331, section 301 of the Federal Food, Drug, and Cosmetic Act. Warning: A willfully and knowingly false statement is a criminal offense, U.S. Code, title 18, section 1001.

11. SIGNATURE OF SPONSOR/APPLICANT/SUBMITTER OR AN AUTHORIZED REPRESENTATIVE (Sign)
Janice Brownlee

12. NAME AND TITLE OF THE PERSON WHO SIGNED IN NO. 11
(Name) Janice Brownlee
(Title) VP, Regulatory Affairs and Quality Assurance

13. ADDRESS *(Number, Street, State, and ZIP Code) (of person identified in Nos. 11 and 12)*
ViewRay Incorporated
2 ThermoFisher Way
Oakwood, OH 44146

14. TELEPHONE AND FAX NUMBERS *(Include Area Code)*
(Tel.) 440-703-3210
(Fax) 440-703-3229

15. DATE OF CERTIFICATION
9/30/10

Form FDA 3674 (11/08) (FRONT)

Declarations of Conformity

ViewRay™ Incorporated commits to conforming with the applicable sections of the following recognized consensus standards for its Treatment Planning and Delivery System, prior to release of the software for commercial, clinical use. Summary reports for each standard are provided.

IEC 61217 Radiotherapy equipment- Coordinates, movements and scales

IEC 62083 Medical electrical equipment: Requirements for the safety of radiotherapy treatment planning systems

IEC 60601-1 Medical electrical equipment-Part1: General Requirements for Safety, 1988; Amendment 1, 1991-11, Amendment 2, 1995

IEC 60601-2-11 Particular Requirements for the safety of gamma beam therapy equipment

Form Approved: OMB No. 0910-0120; Expiration Date: 8/31/10

Department of Health and Human Services
Food and Drug Administration

STANDARDS DATA REPORT FOR 510(k)s

(To be filled in by applicant)

This report and the Summary Report Table are to be completed by the applicant when submitting a 510(k) that references a national or international standard. A separate report is required for each standard referenced in the 510(k).

TYPE OF 510(K) SUBMISSION

☑ Traditional ☐ Special ☐ Abbreviated

STANDARD TITLE [1]

Evaluation of IEC 61217 Ed. 1.2 b (2008) Radiotherapy equipment – Coordinates, movements and scales

Please answer the following questions	*Yes*	*No*
Is this standard recognized by FDA[2]?	☑	☐
FDA Recognition number[3]	# 12-190	
Was a third party laboratory responsible for testing conformity of the device to this standard identified in the 510(k)?	☐	☑
Is a summary report[4] describing the extent of conformance of the standard used included in the 510(k)? If no, complete a summary report table.	☑	☐
Does the test data for this device demonstrate conformity to the requirements of this standard as it pertains to this device?	☑	☐
Does this standard include acceptance criteria? If no, include the results of testing in the 510(k).	☑	☐
Does this standard include more than one option or selection of tests? If yes, report options selected in the summary report table.	☐	☑
Were there any deviations or adaptations made in the use of the standard?	☐	☑
If yes, were deviations in accordance with the FDA supplemental information sheet (SIS)[5]?	☐	☐
Were deviations or adaptations made beyond what is specified in the FDA SIS? If yes, report these deviations or adaptations in the summary report table.	☐	☑
Were there any exclusions from the standard? If yes, report these exclusions in the summary report table.	☑	☐
Is there an FDA guidance[6] that is associated with this standard?	☐	☑
If yes, was the guidance document followed in preparation of this 510k?	☐	☐

Title of guidance: There is no relevant guidance developed at this time.

[1] The formatting convention for the title is: [SDO] [numeric identifier] [title of standard] [date of publication]

[2] Authority [21 U.S.C. 360d], www.fda.gov/cdrh/stdsprog.html

[3] http://www.accessdata.fda.gov/scripts/cdrh/cfdocs/cfStandards/search.cfm

[4] The summary report should include: any adaptations used to adapt to the device under review (for example, alternative test methods); choices made when options or a selection of methods are described; deviations from the standard; requirements not applicable to the device; and the name and address of the test laboratory or certification body involved in conformance assessment to this standard. The summary report includes information on all standards utilized during the development of the device.

[5] The supplemental information sheet (SIS) is additional information which is necessary before FDA recognizes the standard. Found at http://www.accessdata.fda.gov/scripts/cdrh/cfdocs/cfStandards/search.cfm

[6] The online search for CDRH Guidance Documents can be found at www.fda.gov/cdrh/guidance.html

EXTENT OF STANDARD CONFORMANCE
SUMMARY REPORT TABLE

STANDARD TITLE

Evaluation of IEC 61217 Ed. 1.2 b (2008) Radiotherapy equipment – Coordinates, movements and scales

CONFORMANCE WITH STANDARD SECTIONS*

SECTION NUMBER	SECTION TITLE	CONFORMANCE?
See attached	See attached table	☐ Yes ☐ No ☐ N/A

TYPE OF DEVIATION OR OPTION SELECTED *

DESCRIPTION

JUSTIFICATION

SECTION NUMBER	SECTION TITLE	CONFORMANCE?
		☐ Yes ☐ No ☐ N/A

TYPE OF DEVIATION OR OPTION SELECTED *

DESCRIPTION

JUSTIFICATION

SECTION NUMBER	SECTION TITLE	CONFORMANCE?
		☐ Yes ☐ No ☐ N/A

TYPE OF DEVIATION OR OPTION SELECTED *

DESCRIPTION

JUSTIFICATION

* For completeness list all sections of the standard and indicate whether conformance is met. If a section is not applicable (N/A) an explanation is needed under "justification." Some standards include options, so similar to deviations, the option chosen needs to be described and adequately justified as appropriate for the subject device. Explanation of all deviations or description of options selected when following a standard is required under "type of deviation or option selected," "description" and "justification" on the report. More than one page may be necessary.

* Types of deviations can include an exclusion of a section in the standard, a deviation brought out by the FDA supplemental information sheet (SIS), a deviation to adapt the standard to the device, or any adaptation of a section.

Paperwork Reduction Act Statement

Public reporting burden for this collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing the collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden, to:

Center for Devices and Radiological Health
1350 Piccard Drive
Rockville, MD 20850

An agency may not conduct or sponsor, and a person is not required to respond to, a collection of information unless it displays a currently valid OMB control number.

IEC 61217 Ed. 1.2 b (2008)

The ViewRay™ Treatment Planning and Delivery System's software user interface has been designed to comply with IEC 61217 with the following exclusions:

Sections	**Reason**
2.1.5, 2.7, 4., 5., 6.2, 6.5, 6.6, 6.7.4, and 6.7.5	Partially applicable—the ViewRay system was not designed to include the need for rotation of table
2.5, 6.3, and 6.8.1-6.8.5	The ViewRay System does not require wedge filters
2.6, 4., 5., and 6.8.1-6.8.5	The ViewRay System does not incorporate an x-ray image receptor. Imaging on the system is done with magnetic resonance not x-rays.
Justification	
Treatment on the ViewRay System does not require table rotation, wedge filters or x-ray image reception; beam shaping is accomplished with a multi-leaf collimator and imaging is done with magnetic resonance not x-rays	

Form Approved: OMB No. 0910-0120; Expiration Date: 8/31/10

Department of Health and Human Services
Food and Drug Administration

STANDARDS DATA REPORT FOR 510(k)s

(To be filled in by applicant)

This report and the Summary Report Table are to be completed by the applicant when submitting a 510(k) that references a national or international standard. A separate report is required for each standard referenced in the 510(k).

TYPE OF 510(K) SUBMISSION

☑ Traditional ☐ Special ☐ Abbreviated

STANDARD TITLE [1]

IEC 60601-1 {ed3.0}b General requirements for basic safety and essential performance

Please answer the following questions	*Yes*	*No*
Is this standard recognized by FDA [2] ?	☑	☐
FDA Recognition number [3]	# 12-210	
Was a third party laboratory responsible for testing conformity of the device to this standard identified in the 510(k)?	☐	☑
Is a summary report [4] describing the extent of conformance of the standard used included in the 510(k)? If no, complete a summary report table.	☑	☐
Does the test data for this device demonstrate conformity to the requirements of this standard as it pertains to this device?	☑	☐
Does this standard include acceptance criteria? If no, include the results of testing in the 510(k).	☑	☐
Does this standard include more than one option or selection of tests? If yes, report options selected in the summary report table.	☐	☑
Were there any deviations or adaptations made in the use of the standard?	☐	☑
If yes, were deviations in accordance with the FDA supplemental information sheet (SIS) [5] ?	☐	☐
Were deviations or adaptations made beyond what is specified in the FDA SIS? If yes, report these deviations or adaptations in the summary report table.	☐	☑
Were there any exclusions from the standard? If yes, report these exclusions in the summary report table.	☑	☐
Is there an FDA guidance [6] that is associated with this standard?	☐	☑
If yes, was the guidance document followed in preparation of this 510k?	☐	☐

Title of guidance: There is no relevant guidance developed at this time.

[1] The formatting convention for the title is: [SDO] [numeric identifier] [title of standard] [date of publication]

[2] Authority [21 U.S.C. 360d], www.fda.gov/cdrh/stdsprog.html

[3] http://www.accessdata.fda.gov/scripts/cdrh/cfdocs/cfStandards/search.cfm

[4] The summary report should include: any adaptations used to adapt to the device under review (for example, alternative test methods); choices made when options or a selection of methods are described; deviations from the standard; requirements not applicable to the device; and the name and address of the test laboratory or certification body involved in conformance assessment to this standard. The summary report includes information on all standards utilized during the development of the device.

[5] The supplemental information sheet (SIS) is additional information which is necessary before FDA recognizes the standard. Found at http://www.accessdata.fda.gov/scripts/cdrh/cfdocs/cfStandards/search.cfm

[6] The online search for CDRH Guidance Documents can be found at www.fda.gov/cdrh/guidance.html

EXTENT OF STANDARD CONFORMANCE
SUMMARY REPORT TABLE

STANDARD TITLE
IEC 60601-1{ed3.0}b General requirements for basic safety and essential performance

CONFORMANCE WITH STANDARD SECTIONS*

SECTION NUMBER	SECTION TITLE	CONFORMANCE?
See attached	See attached table	☐ Yes ☐ No ☐ N/A

TYPE OF DEVIATION OR OPTION SELECTED *

DESCRIPTION

JUSTIFICATION

SECTION NUMBER	SECTION TITLE	CONFORMANCE?
		☐ Yes ☐ No ☐ N/A

TYPE OF DEVIATION OR OPTION SELECTED *

DESCRIPTION

JUSTIFICATION

SECTION NUMBER	SECTION TITLE	CONFORMANCE?
		☐ Yes ☐ No ☐ N/A

TYPE OF DEVIATION OR OPTION SELECTED *

DESCRIPTION

JUSTIFICATION

* For completeness list all sections of the standard and indicate whether conformance is met. If a section is not applicable (N/A) an explanation is needed under "justification." Some standards include options, so similar to deviations, the option chosen needs to be described and adequately justified as appropriate for the subject device. Explanation of all deviations or description of options selected when following a standard is required under "type of deviation or option selected," "description" and "justification" on the report. More than one page may be necessary.

* Types of deviations can include an exclusion of a section in the standard, a deviation brought out by the FDA supplemental information sheet (SIS), a deviation to adapt the standard to the device, or any adaptation of a section.

Paperwork Reduction Act Statement

Public reporting burden for this collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing the collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden, to:

Center for Devices and Radiological Health
1350 Piccard Drive
Rockville, MD 20850

An agency may not conduct or sponsor, and a person is not required to respond to, a collection of information unless it displays a currently valid OMB control number.

IEC 60601-1{ed3.0}b

The ViewRay™ Treatment Planning and Delivery System's software user interface has been designed to comply with IEC 60601-1 with the following exclusions:

Sections	**Reason**
4.4, 7.9.2.1, 7.9.2.9,	ViewRay Inc. shall provide instructions for use document to the customer.
7.9.2.16	ViewRay Inc. shall provide service personnel a test plan with each maintenance update.
4.6, 4.8, 4.9, 4.10.1, 4.10.2, 5.7, 5.9.2.1, 5.9.2.2, 6.1-6.6, 7.1.2, 7.1.3, 7.2.1, 7.2.4-7.2.7, 7.2.8.1, 7.2.8.2, 7.2.9-7.2.20, 7.3.1-7.3.8, 7.4.1, 7.4.2, 7.5, 7.6.2, 7.6.3, 7.7.1-7.7.5, 7.9.2.2-7.9.2.8, 7.9.2.11-7.9.2.15, 7.9.3.1-7.9.3.4, Section 8, Section 9, Section 10, Section 11, Section 12, 13, 14.8, Section 15, Section 16, 17.	The standards requirement does not pertain to software.
5.9.1	Documents
5.9.2.3	Mechanical
Justification	
ViewRay Inc. shall provide instructions for use document to the customer. ViewRay Inc. shall provide service personnel a test plan with each maintenance update. The standards requirement does not pertain to software.	

Form Approved: OMB No. 0910-0120; Expiration Date: 8/31/10

Department of Health and Human Services
Food and Drug Administration

STANDARDS DATA REPORT FOR 510(k)s

(To be filled in by applicant)

This report and the Summary Report Table are to be completed by the applicant when submitting a 510(k) that references a national or international standard. A separate report is required for each standard referenced in the 510(k).

TYPE OF 510(K) SUBMISSION

☑ Traditional ☐ Special ☐ Abbreviated

STANDARD TITLE [1]

Evaluation of IEC 60601-2-11{ed2.0}b gamma beam therapy equipment

Please answer the following questions	*Yes*	*No*
Is this standard recognized by FDA[2]?	☑	☐
FDA Recognition number[3]	# 12-133	
Was a third party laboratory responsible for testing conformity of the device to this standard identified in the 510(k)?	☐	☑
Is a summary report[4] describing the extent of conformance of the standard used included in the 510(k)? If no, complete a summary report table.	☑	☐
Does the test data for this device demonstrate conformity to the requirements of this standard as it pertains to this device?	☑	☐
Does this standard include acceptance criteria? If no, include the results of testing in the 510(k).	☑	☐
Does this standard include more than one option or selection of tests? If yes, report options selected in the summary report table.	☐	☑
Were there any deviations or adaptations made in the use of the standard?	☐	☑
If yes, were deviations in accordance with the FDA supplemental information sheet (SIS)[5]?	☐	☐
Were deviations or adaptations made beyond what is specified in the FDA SIS? If yes, report these deviations or adaptations in the summary report table.	☐	☑
Were there any exclusions from the standard? If yes, report these exclusions in the summary report table.	☑	☐
Is there an FDA guidance[6] that is associated with this standard?	☐	☑
If yes, was the guidance document followed in preparation of this 510k?	☐	☐

Title of guidance: There is no relevant guidance developed at this time.

[1] The formatting convention for the title is: [SDO] [numeric identifier] [title of standard] [date of publication]

[2] Authority [21 U.S.C. 360d], www.fda.gov/cdrh/stdsprog.html

[3] http://www.accessdata.fda.gov/scripts/cdrh/cfdocs/cfStandards/search.cfm

[4] The summary report should include: any adaptations used to adapt to the device under review (for example, alternative test methods); choices made when options or a selection of methods are described; deviations from the standard; requirements not applicable to the device; and the name and address of the test laboratory or certification body involved in conformance assessment to this standard. The summary report includes information on all standards utilized during the development of the device.

[5] The supplemental information sheet (SIS) is additional information which is necessary before FDA recognizes the standard. Found at http://www.accessdata.fda.gov/scripts/cdrh/cfdocs/cfStandards/search.cfm

[6] The online search for CDRH Guidance Documents can be found at www.fda.gov/cdrh/guidance.html

EXTENT OF STANDARD CONFORMANCE
SUMMARY REPORT TABLE

STANDARD TITLE
Evaluation of IEC 60601-2-11{ed2.0}b gamma beam therapy equipment

CONFORMANCE WITH STANDARD SECTIONS*

SECTION NUMBER	SECTION TITLE	CONFORMANCE?
See attached	See attached table	☐ Yes ☐ No ☐ N/A

TYPE OF DEVIATION OR OPTION SELECTED *

DESCRIPTION

JUSTIFICATION

SECTION NUMBER	SECTION TITLE	CONFORMANCE?
		☐ Yes ☐ No ☐ N/A

TYPE OF DEVIATION OR OPTION SELECTED *

DESCRIPTION

JUSTIFICATION

SECTION NUMBER	SECTION TITLE	CONFORMANCE?
		☐ Yes ☐ No ☐ N/A

TYPE OF DEVIATION OR OPTION SELECTED *

DESCRIPTION

JUSTIFICATION

* For completeness list all sections of the standard and indicate whether conformance is met. If a section is not applicable (N/A) an explanation is needed under "justification." Some standards include options, so similar to deviations, the option chosen needs to be described and adequately justified as appropriate for the subject device. Explanation of all deviations or description of options selected when following a standard is required under "type of deviation or option selected," "description" and "justification" on the report. More than one page may be necessary.

* Types of deviations can include an exclusion of a section in the standard, a deviation brought out by the FDA supplemental information sheet (SIS), a deviation to adapt the standard to the device, or any adaptation of a section.

Paperwork Reduction Act Statement

Public reporting burden for this collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing the collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden, to:

Center for Devices and Radiological Health
1350 Piccard Drive
Rockville, MD 20850

An agency may not conduct or sponsor, and a person is not required to respond to, a collection of information unless it displays a currently valid OMB control number.

IEC 60601-2-11{ed2.0}b and [IEC60601-2-11-amd1{ed2.0}en

The ViewRay™ Treatment Planning and Delivery System's software user interface has been designed to comply with IEC 60601-2-11 with the following exclusions:

Sections	**Reason**
4.1, 4.6, 4.8, 4.10, 5.1-5.6, 6.1-6.3, 6.7, 6.8.1-6.8.3, 10.1, 10.2, 16., 18., 19.-19.4, 20., 21., 22.4, 27., 28., 29.1.1.1-29.1.1.4, 29.1.3.3, 29.1.3.5, 29.1.3.6, 29.1.5.1, 29.1.5.2, 29.1.6.1-29.1.6.3, 29.1.7.1, 29.1.7.2, 29.1.8-29.1.10, 29-1-12, 29.3.1.1, 29.3.1.2, 29.3.2, 29.4.1.1, 29.4.2, 29.4.3.1-29.4.3.3, 29.4.4.1-29.4.4.5, 29.4.5, 29.4.6, 57.1	The standards requirement does not pertain to software.
29.	ViewRay Inc. shall provide an instructions for use document to the customer.
29.1.1.1	The control system shall provide the means to return the source to its off position.
29.1.1.2	The control system shall ensure timing of all source movements.
29.1.4.1, 29.1.4.2	The ViewRay™ System is only capable of stationary therapy.
29.2.1	The ViewRay™ System shall allow for QA plans to be executed to verify absorbed dose.
Justification	
The standards requirement does not pertain to software. ViewRay Inc. shall provide an instructions for use document to the customer. The control system shall provide the means to return the source to its off position. The control system shall ensure timing of all source movements. The ViewRay™ System is only capable of stationary therapy. The ViewRay™ System shall allow for QA plans to be executed to verify absorbed dose.	

Form Approved: OMB No. 0910-0120; Expiration Date: 8/31/10

Department of Health and Human Services
Food and Drug Administration

STANDARDS DATA REPORT FOR 510(k)s

(To be filled in by applicant)

This report and the Summary Report Table are to be completed by the applicant when submitting a 510(k) that references a national or international standard. A separate report is required for each standard referenced in the 510(k).

TYPE OF 510(K) SUBMISSION

☑ Traditional ☐ Special ☐ Abbreviated

STANDARD TITLE [1]

IEC 62083{ed1.0}b Requirements for the safety of radiotherapy treatment planning systems

Please answer the following questions	*Yes*	*No*
Is this standard recognized by FDA [2]?	☑	☐
FDA Recognition number [3] # 12-122		
Was a third party laboratory responsible for testing conformity of the device to this standard identified in the 510(k)?	☐	☑
Is a summary report [4] describing the extent of conformance of the standard used included in the 510(k)? If no, complete a summary report table.	☑	☐
Does the test data for this device demonstrate conformity to the requirements of this standard as it pertains to this device?	☑	☐
Does this standard include acceptance criteria? If no, include the results of testing in the 510(k).	☑	☐
Does this standard include more than one option or selection of tests? If yes, report options selected in the summary report table.	☐	☑
Were there any deviations or adaptations made in the use of the standard?	☐	☑
If yes, were deviations in accordance with the FDA supplemental information sheet (SIS) [5]?	☐	☐
Were deviations or adaptations made beyond what is specified in the FDA SIS? If yes, report these deviations or adaptations in the summary report table.	☐	☑
Were there any exclusions from the standard? If yes, report these exclusions in the summary report table.	☑	☐
Is there an FDA guidance [6] that is associated with this standard?	☐	☑
If yes, was the guidance document followed in preparation of this 510k?	☐	☐

Title of guidance: There is no relevant guidance developed at this time.

[1] The formatting convention for the title is: [SDO] [numeric identifier] [title of standard] [date of publication]

[2] Authority (21 U.S.C. 360d), www.fda.gov/cdrh/stdsprog.html

[3] http://www.accessdata.fda.gov/scripts/cdrh/cfdocs/cfStandards/search.cfm

[4] The summary report should include: any adaptations used to adapt to the device under review (for example, alternative test methods); choices made when options or a selection of methods are described; deviations from the standard; requirements not applicable to the device; and the name and address of the test laboratory or certification body involved in conformance assessment to this standard. The summary report includes information on all standards utilized during the development of the device.

[5] The supplemental information sheet (SIS) is additional information which is necessary before FDA recognizes the standard. Found at http://www.accessdata.fda.gov/scripts/cdrh/cfdocs/cfStandards/search.cfm

[6] The online search for CDRH Guidance Documents can be found at www.fda.gov/cdrh/guidance.html

EXTENT OF STANDARD CONFORMANCE
SUMMARY REPORT TABLE

STANDARD TITLE

IEC 62083{ed1.0}b Requirements for the safety of radiotherapy treatment planning systems

CONFORMANCE WITH STANDARD SECTIONS*

SECTION NUMBER	SECTION TITLE	CONFORMANCE?
See attached	See attached table	☐ Yes ☐ No ☐ N/A

TYPE OF DEVIATION OR OPTION SELECTED *

DESCRIPTION

JUSTIFICATION

SECTION NUMBER	SECTION TITLE	CONFORMANCE?
		☐ Yes ☐ No ☐ N/A

TYPE OF DEVIATION OR OPTION SELECTED *

DESCRIPTION

JUSTIFICATION

SECTION NUMBER	SECTION TITLE	CONFORMANCE?
		☐ Yes ☐ No ☐ N/A

TYPE OF DEVIATION OR OPTION SELECTED *

DESCRIPTION

JUSTIFICATION

* For completeness list all sections of the standard and indicate whether conformance is met. If a section is not applicable (N/A) an explanation is needed under "justification." Some standards include options, so similar to deviations, the option chosen needs to be described and adequately justified as appropriate for the subject device. Explanation of all deviations or description of options selected when following a standard is required under "type of deviation or option selected," "description" and "justification" on the report. More than one page may be necessary.

• Types of deviations can include an exclusion of a section in the standard, a deviation brought out by the FDA supplemental information sheet (SIS), a deviation to adapt the standard to the device, or any adaptation of a section.

Paperwork Reduction Act Statement

Public reporting burden for this collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing the collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden, to:

Center for Devices and Radiological Health
1350 Piccard Drive
Rockville, MD 20850

An agency may not conduct or sponsor, and a person is not required to respond to, a collection of information unless it displays a currently valid OMB control number.

IEC 62083{ed1.0}b

The ViewRay™ Treatment Planning and Delivery System's software user interface has been designed to comply with IEC 62083 with the following exclusions:

Sections	Reason
3.1, 5.2	Hardware
6.	Technical Description
7.7	1. DICOM® is used for images, dose, contours, and points import and export. 2. TCP/IP over Ethernet is used for communications between computer systems. Has inherent error checking.
13.	POST test done by computer hardware. Computer comes with diagnostic tools in the BIOS program.
14.	There is no arithmetic co-processor in the processor being used.
Justification	
DICOM is used for images, dose, contours, and points import and export. TCP/IP over Ethernet is used for communications between computer systems. Has inherent error checking. POST test done by computer hardware. Computer comes with diagnostic tools in the BIOS program. There is no arithmetic co-processor in the processor being used. Hardware Technical Description	

DICOM® is the registered trademark of the National Electrical Manufacturers Association for its standards publications relating to digital communications of medical information.

Executive Summary for ViewRay™ Treatment Planning and Delivery System Software (TPDS)

The ViewRay™ Treatment Planning and Delivery System software is intended to be used for planning and delivering external beam radiation treatments with the ViewRay radiation therapy system. Use of the MRI-guided radiation therapy delivery system and software is indicated for stereotactic radiosurgery and precision radiotherapy for malignant or benign lesions, tumors, and conditions anywhere in the body when radiation treatment is indicated.

The ViewRay System is an MR image-guided radiation therapy device. Similar to other existing image-guided radiation therapy (IGRT) systems in the marketplace, the ViewRay™ System combines two existing technologies: (1) a radiological imaging device for patient positioning and target monitoring and (2) a radiation therapy planning and delivery system. As requested by the FDA, this initial submission installment covers the treatment planning and delivery system software (TPDS) of the device. A subsequent submission installment will cover the hardware aspects of the ViewRay™ System and its integration with the TPDS.

The Varian Medical Systems' Trilogy®/Eclipse® System is an integrated system that includes treatment-planning, imaging, and treatment delivery functions that are directly analogous to the functions provided by the ViewRay™ System. Both the Varian and ViewRay™ Systems can use images obtained from CT, PET or MRI for planning. Although the Varian system uses a different technology for obtaining images during delivery of treatment (CT vs. MRI) and as a source of radiation (Linac vs. Cobalt-60), both systems have the same intended use and indications for use, and are used by the same user population. The imaging and therapy technologies in the Varian system, CT and linear acceleration, were also well established in the marketplace prior to being combined into an integrated IGRT system. In the same way, the ViewRay™ System is integrating two well-established technologies—MRI and Cobalt 60—with treatment planning to provide an integrated solution.

The ViewRay™ Treatment Planning and Delivery System software provides tools for planning and delivery of external gamma beam stereotactic radiosurgery and precision radiotherapy for lesions, tumors, and conditions anywhere in the body when radiation treatment is indicated. It is a computer-based software device used only on the ViewRay™ image-guided radiation therapy system by trained medical professionals. The ViewRay™ TPDS is capable of assisting clinicians in reviewing, prescribing, tracking, and correcting the course of patient treatment using tools for contouring, visualization, data storage, anatomical target monitoring and re-optimization.

This premarketing notification describes the ViewRay™ Treatment Planning and Delivery System software (TPDS) and discusses how its performance outputs are equivalent to those of the cited predicate, Varian Medical Systems Trilogy®/Eclipse® System.

The intended use, indications for use, and user population are the same for both software programs. The ViewRay™ Treatment Planning and Delivery System software also shares many of the technological features and characteristics of the Varian Eclipse planning system, and the dose calculation and treatment delivery features of the Varian Trilogy System software. The fundamental technical characteristics are the same as those of the predicate device and differences are described in the comparison chart and discussion provided in Section 3 of this 510(k) submission.

Verification of the ViewRay™ Treatment Planning and Delivery System software has been thoroughly conducted to demonstrate that the software meets its requirements and intended use. 221 tests were executed to verify the software product requirements, hazard mitigations, use cases and applicable standards. In addition, users including a practicing radiation oncologist, dosimetrist, radiation therapist and physicist who will use the system software were engaged in various protocols to validate that the functions of the system would meet their needs for radiation therapy prescribing, planning, QA, and delivery. The results of these tests are presented in this submission and support ViewRay™ software's ability to meet its intended use and its equivalence to its predicates.

As demonstrated in the enclosed discussions and documentation, we believe the ViewRay™ System TPDS is substantially equivalent to the Varian Trilogy/Eclipse System software for the intended use of treatment planning and delivery of image-guided radiation therapy. The predicate device comparison chart is presented below and discussed in Section 3 of this submission.

Parameter	ViewRay ™TPDS	Varian Eclipse™ Treatment Planning System K091492, FDA clearance date 6/18/09	Varian Trilogy Mx™ System K092871, FDA clearance date 11/30/09 (TrueBeam)
Intended use	The ViewRay Treatment Planning and Delivery System Software is intended to be used for planning external beam irradiation with photon beams and delivering stereotactic radiosurgery and precision radiotherapy for lesions, tumors, and conditions anywhere in the body when radiation treatment is indicated, in conjunction with the ViewRay System, an MRI image-guided radiation therapy system.	The Eclipse Treatment Planning System (Eclipse TPS) is used to plan radiotherapy treatments for patients with malignant or benign diseases. Eclipse TPS is used to plan external beam irradiation with photon, electron and proton beams, as well as for internal irradiation (brachytherapy) treatments. In addition, the Eclipse Proton Eye algorithm is specifically indicated for planning proton treatment of neoplasms of the eye.	The Varian Trilogy is intended to provide stereotactic radiosurgery and precision radiotherapy for lesions, tumors, and conditions anywhere in the body where radiation treatment is indicated.
Indications for use	Use of the ViewRay Treatment Planning and Delivery Software is indicated for stereotactic radiosurgery and precision radiotherapy for lesions, tumors, and conditions anywhere in the body when radiation treatment is indicated.	The Eclipse Treatment Planning System (Eclipse TPS) is used to plan radiotherapy treatments for patients with malignant or benign diseases. Eclipse TPS is used to plan external beam irradiation with photon, electron and proton beams, as well as for internal irradiation (brachytherapy) treatments. In addition, the Eclipse Proton Eye algorithm is specifically indicated for planning proton treatment of neoplasms of the eye.	The Varian Trilogy is intended to provide stereotactic radiosurgery and precision radiotherapy for lesions, tumors, and conditions anywhere in the body where radiation treatment is indicated.

Parameter	ViewRay ™TPDS	Varian Eclipse™ Treatment Planning System K091492, FDA clearance date 6/18/09	Varian Trilogy Mx™ System K092871, FDA clearance date 11/30/09 (TrueBeam)
Classification	21CFR 892.5750 Radionuclide Radiation Therapy System, Class II, Product codes: IWB	21 CFR 892.5050, Medical Charged-particle Radiation Therapy System, Class II Product Codes: MUJ	21 CFR 892.5050, Medical Charged-particle Radiation Therapy System, Class II Product Codes: IYE
User population	Radiation Oncology Depts./Facilities	Radiation Oncology Depts./Facilities	Radiation Oncology Depts./Facilities
Design	Software for treatment planning and MRI image-guided radiotherapy delivery	Software for treatment planning	Software for kV image-guided radiotherapy delivery

TREATMENT PLANNING COMPARISON:

Feature	ViewRay ™TPDS	Varian Eclipse™ Treatment Planning System K091492, FDA clearance date 6/18/09
[redacted]	[redacted]	[redacted] (b)(4)

Feature	ViewRay ™TPDS	Varian Eclipse™ Treatment Planning System K091492, FDA clearance date 6/18/09
(b)(4)		

Feature	ViewRay ™TPDS	Varian Eclipse™ Treatment Planning System K091492, FDA clearance date 6/18/09
(b)(4)		

TREATMENT DELIVERY COMPARISON:

Feature	ViewRay ™TPDS	Varian Trilogy Mx™ System K092871, FDA clearance date 11/30/09 (TrueBeam)
(b)(4)		

Feature	ViewRay ™TPDS	Varian Trilogy Mx™ System K092871, FDA clearance date 11/30/09 (TrueBeam)
(b)(4)		

Section 2

DESCRIPTIVE INFORMATION for the VIEWRAY™ TREATMENT PLANNING and DELIVERY SYSTEM SOFTWARE

(b)(4), Software Design, Software Requirements

(b)(4), Software Design, Software Requirements

(b)(4), Software Design, Software Requirements

(b)(4), Software Design, Software Requirements

(b)(4), Software Design, Software Requirements

(b)(4), Software Design, Software Requirements

(b)(4), Software Design, Software Requirements

(b)(4), Software Design, Software Requirements

(b)(4), Software Design, Software Requirements

(b)(4), Software Design, Software Requirements

(b)(4), Software Design, Software Requirements

(b)(4), Software Design, Software Requirements

(b)(4), Software Design, Software Requirements

(b)(4), Software Design, Software Requirements

(b)(4), Software Design, Software Requirements

(b)(4), Software Design, Software Requirements

(b)(4), Software Design, Software Requirements

(b)(4), Software Design, Software Requirements

(b)(4), Software Design, Software Requirements

(b)(4), Software Design, Software Requirements

(b)(4), Software Design, Software Requirements

(b)(4), Software Design, Software Requirements

(b)(4), Software Design, Software Requirements

(b)(4), Software Design, Software Requirements

(b)(4), Software Design, Software Requirements

(b)(4), Software Design, Software Requirements

(b)(4), Software Design, Software Requirements

(b)(4), Software Design, Software Requirements

Section 3

SUBSTANTIAL EQUIVALENCE DISCUSSION for the VIEWRAY™ TREATMENT PLANNING and DELIVERY SYSTEM

3A. Existing Predicate Devices

ViewRay™ has selected predicate devices based on the intended and indicated use of their image guided radiation therapy (IGRT) system, as discussed with the FDA in previous interactions. The ViewRay™ Treatment Planning System predicate device is the Varian Medical Systems Eclipse™ Treatment Planning System. The ViewRay™ predicate device for treatment delivery is the Varian Trilogy Mx (TrueBeam™) System.

The ViewRay™ Treatment Planning and Delivery Software is intended to be used for planning and delivering stereotactic radiosurgery and precision radiotherapy for lesions, tumors, and conditions anywhere in the body when radiation treatment is indicated, in conjunction with the ViewRay™ System, an MRI image-guided radiation therapy system.

Use of the ViewRay™ Treatment Planning and Delivery Software is indicated for stereotactic radiosurgery and precision radiotherapy for lesions, tumors, and conditions anywhere in the body when radiation treatment is indicated.

Currently there are multiple IGRT systems in the market, including: the Varian Medical Systems, Inc. Trilogy™Mx (TrueBeam™), the TomoTherapy® Incorporated, Hi-ART®, and the Elekta Synergy®. These IGRT systems utilize X-Ray computed tomography (CT) and/or planar fluoroscopic imaging for patient set-up and target tracking (often requiring an invasive procedure to implant radio opaque markers in the patient anatomy). These systems have either an integrated treatment planning system, or a separate compatible treatment planning system. Varian Medical Systems Trilogy™ MX System, K092871, FDA clearance date November 30, 2009 has also added TrueBeam™ technology, which is stated on Varian's website to "treat a moving target with unprecedented speed and accuracy". The treatment planning software Eclipse is "fully integrated with TrueBeam" according to the Varian TrueBeam website. (See Predicate Labeling Attachments 6B)

In an early 510(k) for the Varian Eclipse treatment planning system summary, K010975, FDA clearance date May 2, 2001, the indications for use stated, "The Varian Eclipse device is used to plan photon and electron radiation therapy treatments employing linear accelerators and other similar [teletherapy] devices with x-ray energies from 1 -50 MeV, as well as Cobalt-60, and electron energies from 1 – 50 MeV. Eclipse will plan 3D radiotherapy treatment approaches to combined modality plans, coplanar and non-coplanar fields, static and ARC fields, beam modifiers, and beam intensity modulators. Eclipse includes also tools for treatment preparation (diagnostic image analysis, contouring and segmentation) and plan review. "

Several subsequent submissions have been done for the Eclipse to add features for treatment planning. The most recent FDA 510(k) clearance for the Eclipse, K091492, was cleared on June 18, 2009 and includes photon treatments in its intended use.

The ViewRay™ Treatment Planning and Delivery System software is very similar to these devices in both planning and delivery of radiotherapy. By contrast, it utilizes magnetic resonance imaging for patient set-up and target monitoring, so that it, too, can deliver treatment while simultaneously accounting for natural motion in the target. Cobalt-60 was chosen as the source of radiation in the ViewRay™ System because it is compatible with MRI technology and has been shown capable of producing high-quality conformal and intensity modulated radiation therapy treatments. Previous investigations important to the ViewRay™ system include Fox C, Romeijn HE, Lynch B, Men C, Aleman DM, Dempsey JF, Comparative analysis of ^{60}Co intensity-modulated radiation therapy. Phys Med Biol. 2008 Jun 21;53(12):3175-88. Epub 2008 May 27; J.L. Schreiner, C.P. Joshi, J. Darko, A. Kerr, G. Salomons, S. Dhanesar. The role of Cobalt-60 in modern radiation therapy: Dose delivery and image guidance. Journal of Medical Physics, Vol. 34, No. 3, (2009) 133-136; and E J Adams and A P Warrington, A comparison between Cobalt-60 and linear accelerator-based treatment plans for conformal and intensity-modulated radiotherapy. The British Journal of Radiology, 81 (2008), 304–310.[14,2,3]

3B. Selected Predicates and Rationale

ViewRay™ Incorporated believes its Treatment Planning and Delivery System software (TPDS) is substantially equivalent to the following predicates: 1) the Varian Medical Trilogy™ Mx (TrueBeam™) and 2) the Varian Eclipse™.

The Varian Medical Systems Trilogy™ Mx/TrueBeam™/Eclipse™ Systems are integrated to provide treatment planning, imaging, and treatment delivery functions. The software functions of the integrated Varian system are directly analogous to the functions provided by the ViewRay™ TPDS System. Both the Varian and ViewRay™ Systems can use images obtained from CT, PET or MRI for planning. Although the Varian system uses a different technology for obtaining images during delivery of treatment (CT vs. MRI) and as a source of radiation (Linac vs. Cobalt-60), both systems have the same intended use and indications for use, and are used by the same user population. The imaging and therapy technologies used in the Varian system, CT and linear accelerator, were well established in the marketplace prior to being combined into an IGRT system. In the same way, the ViewRay™ System is combining two well-established technologies—MRI and radiotherapy delivery using Cobalt 60—with treatment planning and delivery functions to provide a comprehensive solution. The ViewRay™ hardware system will be described in a subsequent submission of information.

Both the radiotherapy planning and delivery predicate devices were classified as Class II devices. The predicate image-guided radiation therapy system and its planning system went through the 510(k) submission review route for marketing clearance. Since this submission is focused on the software aspects of the ViewRay™ System, the predicates were chosen to address those software functions, which share the same intended use as the ViewRay™ TPDS.

3C. Predicate Device Comparison Chart

Parameter	ViewRay™ TPDS	Varian Eclipse™ Treatment Planning System K091492, FDA clearance date 6/18/09	Varian Trilogy Mx™ System K092871, FDA clearance date 11/30/09 (TrueBeam)
Intended use	The ViewRay™ Treatment Planning and Delivery System is intended to be used for planning external beam irradiation with photon beams and delivering stereotactic radiosurgery and precision radiotherapy for lesions, tumors, and conditions anywhere in the body when radiation treatment is indicated, in conjunction with the ViewRay™ System, an MRI image-guided radiation therapy system.	The Eclipse Treatment Planning System (Eclipse TPS) is used to plan radiotherapy treatments for patients with malignant or benign diseases. Eclipse TPS is used to plan external beam irradiation with photon, electron and proton beams, as well as for internal irradiation (brachytherapy) treatments. In addition, the Eclipse Proton Eye algorithm is specifically indicated for planning proton treatment of neoplasms of the eye.	The Varian Trilogy is intended to provide stereotactic radiosurgery and precision radiotherapy for lesions, tumors, and conditions anywhere in the body where radiation treatment is indicated.

Parameter	ViewRay™ TPDS	Varian Eclipse™ Treatment Planning System K091492, FDA clearance date 6/18/09	Varian Trilogy Mx™ System K092871, FDA clearance date 11/30/09 (TrueBeam)
Indications for use	Use of the ViewRay™ Treatment Planning and Delivery System is indicated for stereotactic radiosurgery and precision radiotherapy for lesions, tumors, and conditions anywhere in the body when radiation treatment is indicated.	The Eclipse Treatment Planning System (Eclipse TPS) is used to plan radiotherapy treatments for patients with malignant or benign diseases. Eclipse TPS is used to plan external beam irradiation with photon, electron and proton beams, as well as for internal irradiation (brachytherapy) treatments. In addition, the Eclipse Proton Eye algorithm is specifically indicated for planning proton treatment of neoplasms of the eye.	The Varian Trilogy is intended to provide stereotactic radiosurgery and precision radiotherapy for lesions, tumors, and conditions anywhere in the body where radiation treatment is indicated.
Classification	21CFR 892.5750 Radionuclide Radiation Therapy System, Class II, Product codes: IWB	21 CFR 892.5050, Medical Charged-particle Radiation Therapy System, Class II Product Codes: MUJ	21 CFR 892.5050, Medical Charged-particle Radiation Therapy System, Class II Product Codes: IYE
User population	Radiation Oncology Depts./Facilities	Radiation Oncology Depts./Facilities	Radiation Oncology Depts./Facilities
Design	Software for treatment planning and MRI image-guided radiotherapy delivery	Software for treatment planning	Software for kV image-guided radiotherapy delivery

TREATMENT PLANNING COMPARISON

Feature	ViewRay™ TPDS	Varian Eclipse™ Treatment Planning System K091492, FDA clearance date 6/18/09
(b)(4)	[redacted]	

Feature	ViewRay™ TPDS	Varian Eclipse™ Treatment Planning System K091492, FDA clearance date 6/18/09
(b)(4)		

Feature	ViewRay™ TPDS	Varian Eclipse™ Treatment Planning System K091492, FDA clearance date 6/18/09
(b)(4)		

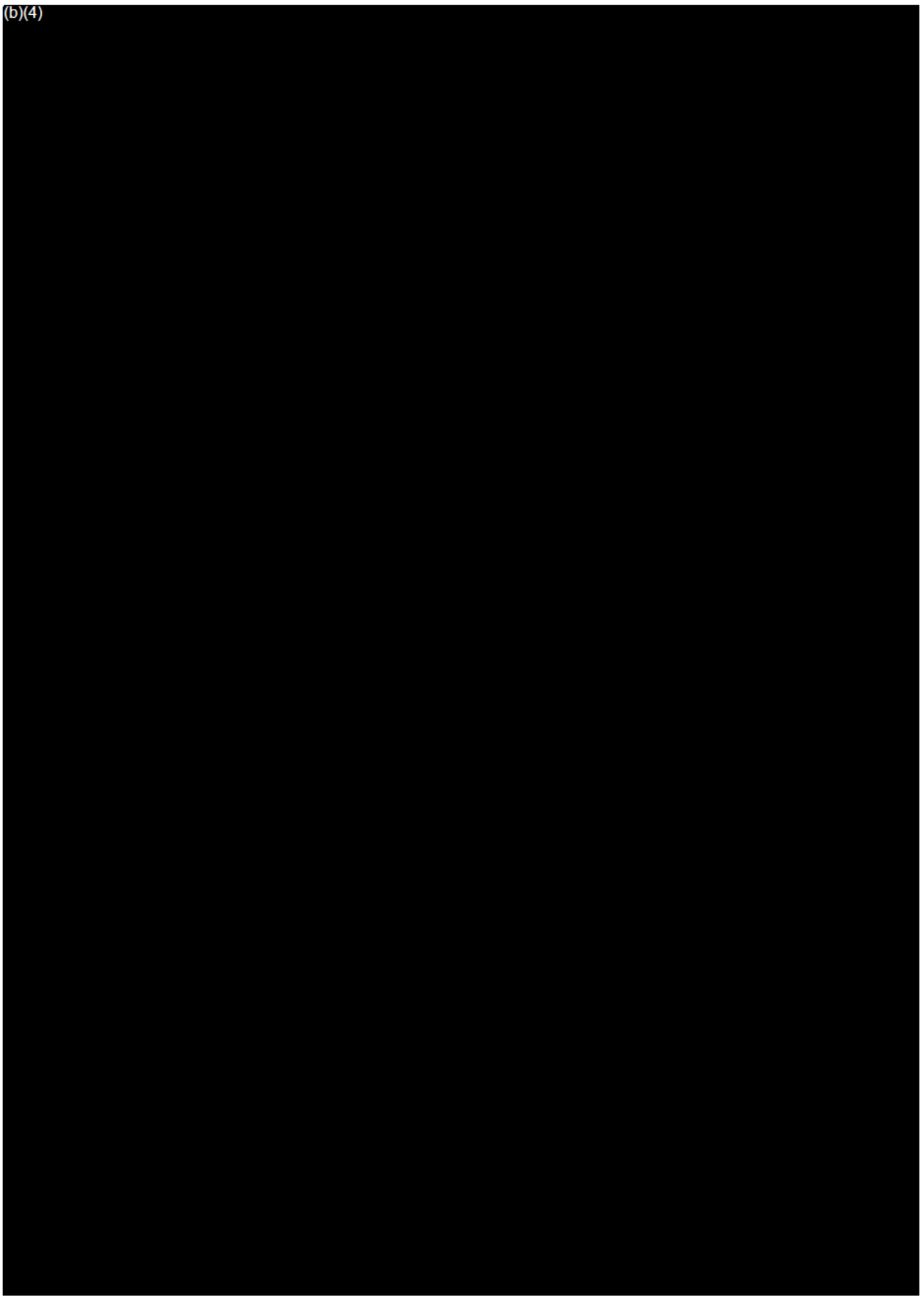
TREATMENT DELIVERY SOFTWARE FEATURES

Feature	ViewRay™ TPDS	Varian Trilogy Mx™ System K092871, FDA clearance date 11/30/09 (TrueBeam™)
(b)(4)		

Feature	ViewRay™ TPDS	Varian Trilogy Mx™ System K092871, FDA clearance date 11/30/09 (TrueBeam™)
(b)(4)		

Feature	ViewRay™ TPDS	Varian Trilogy Mx™ System K092871, FDA clearance date 11/30/09 (TrueBeam™)
(b)(4)		

(b)(4)



The Varian Trilogy software is an integrated part of the Trilogy System for radiation treatment, just as the ViewRay™ treatment delivery system software will be an integrated part of the ViewRay™ Treatment Planning and Delivery System. The Varian Trilogy software, like the ViewRay™ treatment delivery software system, is intended to treat benign or malignant conditions, throughout the body, where radiation therapy is indicated. The software in both systems is intended for use only by trained professionals according to an approved treatment plan.

Delivery gating capability*

Both the Varian and ViewRay™ software have motion management visualization software capabilities to enable treatment of moving tumors, and interrupt treatment if the targeted area deviates from preset parameters. The Varian system achieves this through the use of surrogate markers, and planar x-ray imaging; while the ViewRay™ system monitors the target position through real-time continuous imaging with MRI, avoiding additional patient exposure to radiation.

On board imaging for volumetric and planar acquisition*

The Varian software relies on radiographic, fluoroscopic, and/or cone beam CT (CBCT) scans for image-guided radiation therapy, which exposes the patient to additional radiation. The ViewRay™ System utilizes planar and volumetric MRI scans for image-guided radiation therapy, which provide no additional dose to the patient and can improve soft tissue contrast. Prior investigations to support this are presented in V.S. Khoo and Joon DL, New developments in MRI for target volume delineation in radiotherapy. Br J Radiol. 2006 Sep; 79 Spec No 1:S2-15[5] and J.N. Brunt, Computed Tomography-Magnetic Resonance Image Registration in Radiotherapy Treatment Planning. Clin Oncol (R Coll Radiol). 2010 Jul 29. [Epub ahead of print]PMID: 20674300.[6]

3E. Selected Predicates and Substantial Equivalence Decision

ViewRay™ believes the treatment planning and delivery software for its ViewRay™ Treatment Planning and Delivery System (TPDS) is substantially equivalent to the Varian Medical Systems Eclipse Treatment Planning system and the software for the Varian Medical Systems' Trilogy Mx™ System.

The ViewRay™ rationale for each of the decision points is presented below.

Does the new device have the same intended use and indication statement?

Yes, the new ViewRay™ device has the same intended use and indication statement and may be "substantially equivalent."

Does the new device have the same technological characteristics, e.g., design, materials, etc.?

The ViewRay™ treatment planning and delivery software, considered alone, does have the same technological characteristics as the software in the predicate device. However, the ViewRay™ Treatment Planning and Delivery System for which it is intended solely to be used, does not have the same technological characteristics as it combines radiation therapy with MRI.

Could the new characteristics affect safety or effectiveness?

Although the ViewRay™ TPDS software itself does not have new technological characteristics compared to predicates, the differing characteristics of the ViewRay™ Treatment Planning and Delivery System (driven by the software) could impact safety as follows:

- the use of MRI will necessitate screening of patients for compatibility with a magnetic field and use of non-ferrous equipment in the treatment room, as is typical for MRI procedures in radiology, but has not been done in a radiotherapy environment
- the use of MRI instead of CT for imaging during treatment will reduce the cumulative unintended radiation administered to the patient during the course of treatment, improving safety.

Do the new characteristics raise new types of safety or effectiveness questions?

No, the differing technological characteristics of the ViewRay™ TPDS to be used with the ViewRay™ software do not raise new types of safety or effectiveness questions. Both MRI for imaging and radiation therapy are well-established technologies whose safety and effectiveness are known, and for which safe handling programs have existed for decades.

Do accepted scientific methods exist for assessing effects of the new characteristics?

Scientific methods for assessing the effects of MRI imaging and Cobalt 60 have existed for decades and include the National Electrical Manufacturers Association (NEMA) and the International Electrotechnical Commission (IEC) standards for MRI, the American Association of Physicists in Medicine (AAPM) and the National Council on Radiation Protection & Measurements (NCRP) recommendations for gamma beam therapy.

Are performance data available to assess effects of new characteristics?

Performance data for these system characteristics will be presented in the subsequent submission for the ViewRay™ TPDS, as they do not apply to this system software submission.

Does the performance data demonstrate equivalence?

Performance data on the ViewRay™ TPDS software planning and delivery functions is presented in the Performance Testing section of this submission and does demonstrate equivalence to the Varian predicates. ViewRay™ therefore believes the ViewRay™ Treatment Planning and Delivery System software is substantially equivalent to the Varian Eclipse and Trilogy software functions.

SECTION 4

RISK CONTROL DISCUSSION for the VIEWRAY™ TREATMENT PLANNING and DELIVERY SYSTEM

(b)(4), Risk Assesment

(b)(4), Risk Assesment

(b)(4), Risk Assesment

(b)(4), Risk Assesment

(b)(4), Risk Assesment

(b)(4), Risk Assesment

(b)(4), Risk Assesment

(b)(4), Risk Assesment

(b)(4), Risk Assesment

(b)(4), Risk Assesment

(b)(4), Risk Assesment

(b)(4), Risk Assesment

(b)(4), Risk Assesment

(b)(4), Risk Assesment

(b)(4), Risk Assesment

(b)(4), Risk Assesment

(b)(4), Risk Assesment

(b)(4), Risk Assesment

(b)(4), Risk Assesment

(b)(4), Risk Assesment

This Page Intentionally Left Blank

This Page Intentionally Left Blank

SECTION 5

LABELING

5A. ViewRay™ Treatment Planning and Delivery System Labeling

The proposed labeling for the ViewRay™ Treatment Planning and Delivery System (TPDS) software includes the Operator's Manual, Attachment 5A-1, and a sample of proposed promotional material as Attachment 5A-2. The software will be installed in the system as part of installation; any software used will be identified with the company name, ViewRay™, and its release number.

The TPDS Operator's Manual for software is intended to be a part of the comprehensive ViewRay™ TPDS manual. Consequently, discussions regarding the layout and operation of the hardware are not included in the manual provided with this part of the submission. A brief overview of the system is provided though, to aid the operator's understanding of the software in the context of the whole device.

The manual is set up according to the workflow for radiation therapy- prescription, planning, review and approval, QA, delivery, record-keeping. The goal of the manual is to provide an easy- to- follow reference for the physicians, dosimetrists, physicists and therapists that facilitates their goal of delivering safe and effective therapy. Use of simple instructions and actual screenshots that accurately reflect what the user will see are provided throughout the manual to achieve this goal.

The manual was used as part of user validation training, and during the actual software validation activities with favorable feedback. The users found the information to be concise, readily located and easily followed to execute the validation protocols. It will be provided in paper form to the users as they have indicated this is the easiest way to have it available for easy reference.

The manual contents were also verified to contain the required and/or recommended statements for applicable regulations, standards, professional task groups, and risk mitigations relative to the software functions.

The sample of promotional material for the ViewRay™ TPDS includes the kinds of statements that would be made regarding the treatment planning and delivery software features. Such statements might be included on the ViewRay™ website, in brochures, product specification sheets, advertisements, posters or presentations.

Labeling is reviewed and controlled by the Regulatory Affairs function prior to distribution to ensure the contents correctly describe the product and its use, meet regulatory requirements and recommendations.

5B. Predicate Labeling

The labeling information for the Varian Eclipse™ and Trilogy Mx™ were obtained from the Varian Inc. website www.varian.com. The Eclipse is their treatment planning system and the Trilogy is their radiation treatment system. Since this submission is describing the ViewRay™ treatment planning and delivery software aspects of the ViewRay™ Treatment Planning and Delivery System, attention should be directed on Varian Eclipse and those features of the Trilogy supported by the software for treatment delivery.

The Varian Eclipse labeling is presented as Attachment 5B-1 and the Varian Trilogy Truebeam™ as Attachment 5B-2.

SECTION 6

PERFORMANCE TESTING for the VIEWRAY™ TREATMENT PLANNING and DELIVERY SYSTEM SOFTWARE

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

(b)(4), Testing, Software, Software Design

VIEWRAY INCORPORATED	Title: **Risk Management Process**	(b)(4), Risk Assessment

(b)(4), Risk Assessment

(b)(4), Risk Assessment

(b)(4), Risk Assessment

(b)(4), Risk Assessment

(b)(4), Risk Assessment

(b)(4), Risk Assessment

(b)(4), Risk Assessment

(b)(4), Risk Assessment

(b)(4), Risk Assessment

(b)(4), Risk Assessment

(b)(4), Risk Assessment

(b)(4), Risk Assessment

(b)(4), Risk Assessment

(b)(4), Risk Assessment

(b)(4), Risk Assessment

(b)(4), Risk Assessment

(b)(4), Risk Assessment

(b)(4), Risk Assessment

(b)(4), Risk Assessment

(b)(4), Risk Assessment

(b)(4), Risk Assessment

(b)(4), Risk Assessment

(b)(4), Risk Assessment

(b)(4), Risk Assessment

(b)(4), Risk Assessment

(b)(4), Risk Assessment

(b)(4), Risk Assessment

(b)(4), Risk Assessment

(b)(4), Risk Assessment

(b)(4), Risk Assessment

(b)(4), Risk Assessment

(b)(4), Risk Assessment

(b)(4), Risk Assessment

(b)(4), Risk Assessment

(b)(4), Risk Assessment

(b)(4), Risk Assessment

(b)(4), Risk Assessment

(b)(4), Risk Assessment

(b)(4), Risk Assessment

(b)(4), Risk Assessment

(b)(4), Risk Assessment

(b)(4), Risk Assessment

(b)(4), Risk Assessment

(b)(4), Risk Assessment

(b)(4), Risk Assessment

(b)(4), Risk Assessment

(b)(4), Risk Assessment

(b)(4), Risk Assessment

(b)(4), Risk Assessment

(b)(4), Risk Assessment

(b)(4), Risk Assessment

(b)(4), Risk Assessment

(b)(4), Risk Assessment

(b)(4), Risk Assessment

(b)(4), Risk Assessment

(b)(4), Risk Assessment

(b)(4), Risk Assessment

(b)(4), Risk Assessment

(b)(4), Risk Assessment

(b)(4), Risk Assessment

(b)(4), Risk Assessment

(b)(4), Risk Assessment

(b)(4), Risk Assessment

(b)(4), Risk Assessment

(b)(4), Risk Assessment

(b)(4), Risk Assessment

(b)(4), Risk Assessment

(b)(4), Risk Assessment

(b)(4), Risk Assessment

(b)(4), Risk Assessment

(b)(4), Risk Assessment

(b)(4), Risk Assessment

(b)(4), Risk Assessment

(b)(4), Risk Assessment

(b)(4), Risk Assessment

(b)(4), Risk Assessment

(b)(4), Risk Assessment

(b)(4), Risk Assessment

(b)(4), Risk Assessment

(b)(4), Risk Assessment

(b)(4), Risk Assessment

(b)(4), Risk Assessment

(b)(4), Risk Assessment

(b)(4), Risk Assessment

(b)(4), Risk Assessment

(b)(4), Risk Assessment

(b)(4), Risk Assessment

(b)(4), Risk Assessment

(b)(4), Risk Assessment

(b)(4), Risk Assessment

(b)(4), Risk Assessment

(b)(4), Risk Assessment

(b)(4), Risk Assessment

(b)(4), Risk Assessment

(b)(4), Risk Assessment

(b)(4), Risk Assessment

(b)(4), Risk Assessment

(b)(4), Risk Assessment

(b)(4), Risk Assessment

(b)(4), Risk Assessment

(b)(4), Risk Assessment

(b)(4), Risk Assessment

(b)(4), Risk Assessment

(b)(4), Risk Assessment

(b)(4), Risk Assessment

(b)(4), Risk Assessment

(b)(4), Risk Assessment

(b)(4), Risk Assessment

(b)(4), Risk Assessment

(b)(4), Risk Assessment

(b)(4), Risk Assessment

(b)(4), Risk Assessment

(b)(4), Risk Assessment

(b)(4), Risk Assessment

(b)(4), Risk Assessment

(b)(4), Risk Assessment

(b)(4), Risk Assessment

(b)(4), Risk Assessment

(b)(4), Risk Assessment

(b)(4), Risk Assessment

(b)(4), Risk Assessment

(b)(4), Risk Assessment

(b)(4), Risk Assessment

(b)(4), Risk Assessment

(b)(4), Risk Assessment

(b)(4), Risk Assessment

(b)(4), Risk Assessment

(b)(4), Risk Assessment

(b)(4), Risk Assessment

(b)(4), Risk Assessment

(b)(4), Risk Assessment

(b)(4), Risk Assessment

(b)(4), Risk Assessment

(b)(4), Risk Assessment

(b)(4), Risk Assessment

(b)(4), Risk Assessment

(b)(4), Risk Assessment

(b)(4), Risk Assessment

(b)(4), Risk Assessment

(b)(4), Risk Assessment

(b)(4), Risk Assessment

(b)(4), Risk Assessment

(b)(4), Risk Assessment

(b)(4), Risk Assessment

(b)(4), Risk Assessment

(b)(4), Risk Assessment

(b)(4), Risk Assessment

(b)(4), Risk Assessment

(b)(4), Risk Assessment

(b)(4), Risk Assessment

(b)(4), Risk Assessment

(b)(4), Risk Assessment

(b)(4), Risk Assessment

(b)(4), Risk Assessment

(b)(4), Risk Assessment

(b)(4), Risk Assessment

(b)(4), Risk Assessment

(b)(4), Risk Assessment

(b)(4), Risk Assessment

(b)(4), Risk Assessment

(b)(4), Risk Assessment

(b)(4), Risk Assessment

(b)(4), Risk Assessment

(b)(4), Risk Assessment

(b)(4), Risk Assessment

(b)(4), Risk Assessment

(b)(4), Risk Assessment

(b)(4), Risk Assessment

(b)(4), Risk Assessment

(b)(4), Risk Assessment

(b)(4), Risk Assessment

(b)(4), Risk Assessment

(b)(4), Risk Assessment

(b)(4), Risk Assessment

(b)(4), Risk Assessment

(b)(4), Risk Assessment

(b)(4), Risk Assessment

(b)(4), Risk Assessment

(b)(4), Risk Assessment

(b)(4), Risk Assessment

(b)(4), Risk Assessment

(b)(4), Risk Assessment

(b)(4), Risk Assessment

(b)(4), Risk Assessment

(b)(4), Risk Assessment

(b)(4), Risk Assessment

(b)(4), Risk Assessment

(b)(4), Risk Assessment

(b)(4), Risk Assessment

(b)(4), Risk Assessment

(b)(4), Risk Assessment

(b)(4), Risk Assessment

(b)(4), Risk Assessment

(b)(4), Risk Assessment

(b)(4), Risk Assessment

(b)(4), Risk Assessment

(b)(4), Risk Assessment

(b)(4), Risk Assessment

(b)(4), Risk Assessment

(b)(4), Risk Assessment

(b)(4), Risk Assessment

(b)(4), Risk Assessment

(b)(4), Risk Assessment

(b)(4), Risk Assessment

(b)(4), Risk Assessment

(b)(4), Risk Assessment

(b)(4), Risk Assessment

(b)(4), Risk Assessment

(b)(4), Risk Assessment

(b)(4), Risk Assessment

(b)(4), Risk Assessment

(b)(4), Risk Assessment

(b)(4), Risk Assessment

(b)(4), Risk Assessment

(b)(4), Risk Assessment

(b)(4), Risk Assessment

(b)(4), Risk Assessment

(b)(4), Risk Assessment

(b)(4), Risk Assessment

(b)(4), Risk Assessment

(b)(4), Risk Assessment

(b)(4), Risk Assessment

(b)(4), Risk Assessment

(b)(4), Risk Assessment

(b)(4), Risk Assessment

(b)(4), Risk Assessment

(b)(4), Risk Assessment

(b)(4), Risk Assessment

(b)(4), Risk Assessment

(b)(4), Risk Assessment

(b)(4), Risk Assessment

(b)(4), Risk Assessment

(b)(4), Risk Assessment

(b)(4), Risk Assessment

(b)(4), Risk Assessment

(b)(4), Risk Assessment

(b)(4), Risk Assessment

(b)(4), Risk Assessment

(b)(4), Risk Assessment

(b)(4), Risk Assessment

(b)(4), Risk Assessment

(b)(4), Risk Assessment

(b)(4), Risk Assessment

(b)(4), Risk Assessment

(b)(4), Risk Assessment

(b)(4), Risk Assessment

(b)(4), Risk Assessment

(b)(4), Risk Assessment

(b)(4), Risk Assessment

(b)(4), Risk Assessment

(b)(4), Risk Assessment

(b)(4), Risk Assessment

(b)(4), Risk Assessment

(b)(4), Risk Assessment

(b)(4), Risk Assessment

(b)(4), Risk Assessment

(b)(4), Risk Assessment

(b)(4), Risk Assessment

(b)(4), Risk Assessment

(b)(4), Risk Assessment

(b)(4), Risk Assessment

(b)(4), Risk Assessment

(b)(4), Risk Assessment

(b)(4), Risk Assessment

(b)(4), Risk Assessment

(b)(4), Risk Assessment

(b)(4), Risk Assessment

(b)(4), Risk Assessment

(b)(4), Risk Assessment

(b)(4), Risk Assessment

(b)(4), Risk Assessment

(b)(4), Risk Assessment

(b)(4), Risk Assessment

(b)(4), Risk Assessment

(b)(4), Risk Assessment

(b)(4), Risk Assessment

(b)(4), Risk Assessment

(b)(4), Risk Assessment

(b)(4), Risk Assessment

(b)(4), Risk Assessment

(b)(4), Risk Assessment

(b)(4), Risk Assessment

(b)(4), Risk Assessment

(b)(4), Risk Assessment

(b)(4), Risk Assessment

(b)(4), Risk Assessment

(b)(4), Risk Assessment

(b)(4), Risk Assessment

(b)(4), Risk Assessment

(b)(4), Risk Assessment

(b)(4), Risk Assessment

(b)(4), Risk Assessment

(b)(4), Risk Assessment

(b)(4), Risk Assessment

(b)(4), Risk Assessment

(b)(4), Risk Assessment

(b)(4), Risk Assessment

(b)(4), Risk Assessment

(b)(4), Risk Assessment

(b)(4), Risk Assessment

(b)(4), Risk Assessment

(b)(4), Risk Assessment

(b)(4), Risk Assessment

(b)(4), Risk Assessment

(b)(4), Risk Assessment

(b)(4), Risk Assessment

(b)(4), Risk Assessment

(b)(4), Risk Assessment

(b)(4), Risk Assessment

(b)(4), Risk Assessment

(b)(4), Risk Assessment

(b)(4), Risk Assessment

(b)(4), Risk Assessment

(b)(4), Risk Assessment

(b)(4), Risk Assessment

(b)(4), Risk Assessment

(b)(4), Risk Assessment

(b)(4), Risk Assessment

(b)(4), Risk Assessment

(b)(4), Risk Assessment

(b)(4), Risk Assessment

(b)(4), Risk Assessment

(b)(4), Risk Assessment

(b)(4), Risk Assessment

(b)(4), Risk Assessment

(b)(4), Risk Assessment

(b)(4), Risk Assessment

(b)(4), Risk Assessment

(b)(4), Risk Assessment

(b)(4), Risk Assessment

(b)(4), Risk Assessment

(b)(4), Risk Assessment

(b)(4), Risk Assessment

(b)(4), Risk Assessment

(b)(4), Risk Assessment

(b)(4), Risk Assessment

(b)(4), Risk Assessment

(b)(4), Risk Assessment

(b)(4), Risk Assessment

(b)(4), Risk Assessment

(b)(4), Risk Assessment

(b)(4), Risk Assessment

(b)(4), Risk Assessment

(b)(4), Risk Assessment

(b)(4), Risk Assessment

(b)(4), Risk Assessment

(b)(4), Risk Assessment

(b)(4), Risk Assessment

(b)(4), Risk Assessment

(b)(4), Risk Assessment

(b)(4), Risk Assessment

(b)(4), Risk Assessment

(b)(4), Risk Assessment

(b)(4), Risk Assessment

(b)(4), Risk Assessment

(b)(4), Risk Assessment

(b)(4), Risk Assessment

(b)(4), Risk Assessment

ViewRay™

Treatment Planning and Delivery System

Operator's Manual

Document No. L-0003

DRAFT DATE: September 29, 2010

THIS PAGE INTENTIONALLY LEFT BLANK

Copyright Statement

ViewRay, Inc., has taken care to ensure the accuracy of this document; however, ViewRay Inc. assumes no liability for errors or omissions and reserves the right to make changes without further notice to any products herein to improve reliability, function or design.

ViewRay provides this manual without warranty of any kind, either implied or expressed, including but not limited to, the implied warranties of merchantability and fitness for a particular purpose.

This document contains proprietary information that is protected by copyright. ViewRay Inc. reserves the right to revise this publication and to make changes in content from time to time without obligation on the part of ViewRay to provide notification of such revision or change.

Trademark Information

ViewRay is a trademark of ViewRay Inc.

Windows® is a registered trademark of Microsoft® Corporation. Other brand or product names are trademarks or registered trademarks of their respective holders.

About this Manual

The ViewRay Treatment Planning and Delivery System is intended for use only with the ViewRay Radiation Therapy System. The instructions for use in this manual pertain only to the use of the ViewRay Treatment Planning and Delivery System and will be incorporated in the comprehensive ViewRay System™ Operator's Manual. The comprehensive system manual will include additional information regarding the operation of the hardware and subsystems, safety and QA procedures and will be part of the hardware submission.

The ViewRay Treatment Planning and Delivery System is for PRESCRIPTION USE ONLY. Federal law restricts the device to sale by or on the order of a physician, or licensed healthcare provider.

Intended Use

The ViewRay™ Treatment Planning and Delivery System is intended to be used for planning external beam irradiation with photon beams and delivering stereotactic radiosurgery and precision radiotherapy for lesions, tumors, and conditions anywhere in the body when radiation treatment is indicated, in conjunction with the ViewRay™ System, an MR image-guided radiation therapy system.

The ViewRay System may be contraindicated for patients with implants due to the presence of the MRI magnetic field. All patients must be screened for MRI compatibility prior to treatment.

NOTE **The ViewRay Treatment Planning and Delivery System must be operated according to the intended use and only by qualified persons with the necessary knowledge in accord with country-specific regulations.**

All personnel must complete user training before operating the system. The user must be familiar with all potential hazards and all safety guidelines. In addition, before using the system, the user must read and understand the operator's manual.

For Service or Questions call ViewRay, Inc. 440.703.3210

CONTENTS

LIST OF TABLES

THIS PAGE INTENTIONALLY LEFT BLANK

LIST OF FIGURES

(Intentionally Blank)

Introduction

About this Manual

This manual provides you with the information needed to properly operate the ViewRay Treatment Planning and Delivery System. The manual contains step-by-step operation instructions for patient registration, treatment planning, on-table planning, anatomical gating, and dose delivery. This manual includes information on the following topics:

- **Introduction**: Provides general information about the ViewRay System and instructions on how to use the manual.
- **Getting Started:** Provides instructions on starting up, logging on and off, switching users, password setup and change.
- **Treatment Planning:** Provides instructions for creating a treatment plan, importing data from DICOM, adding and registering images, optimizing the dose prescription, setting gating parameters, and instructions for using all the ViewRay System™ contouring features.
- **Prescription and Patient Management:** Provides instructions for the basic tasks of registering a patient, viewing, approving, disapproving, and exporting patient data, patient positioning, imaging, setting up rapid treatment, and procedures for fault modes.
- **Treatment Delivery:** Provides instructions for delivering a treatment plan and for using on-table treatment planning which includes contouring, dose prediction, re-optimization, plan approval, and delivery of radiation treatment.

System Overview

The ViewRay System™ is an Image-Guided Radiation Therapy (IGRT) system that uses a Magnetic Resonance Imaging (MRI) unit for image guidance and a three-headed Cobalt-60 radiotherapy gantry for radiation delivery. The ViewRay System uses its integrated 0.35 Tesla MR system (which is capable of acquiring images of high spatial integrity) to track the target during radiation therapy and to pause radiation whenever the target is not within range. This image-guided targeted delivery of radiation is designed to maximize dose to the target while minimizing dose to healthy tissue.

The system is designed so that the imaging and radiotherapy fields of view coincide, making it possible to image the patient at the radiotherapy isocenter before and during treatment. The ViewRay System can use the image data from pretreatment MRI scans to create accurate radiotherapy treatment plans. The system can also image the patient during treatment and based on the constraints and parameters set by a physician, deliver therapy to the dynamic (moving) patient, pausing treatment whenever the target is out of range.

Key Features of the ViewRay System™

The ViewRay Treatment Planning and Delivery System, also referred to as the ViewRay System and ViewRay TPDS, combines MR imaging with fractionated high-precision radiotherapy. The ViewRay System enables real-time, continuous imaging during treatment to ensure accurate radiation delivery. Because the imaging uses MR technology it does not add an additional ionizing radiation dose to the patient.

The ViewRay System is designed to create optimized Intensity-Modulated Radiation Therapy (IMRT), conformal, and combinations of IMRT and conformal treatment plans for delivery only on the ViewRay System. The ViewRay TPDS can aid clinicians in the review, prescription, tracking and correcting of patient treatment.

The following table lists the symbols and formatting conventions used in this manual.

TABLE 1-1. Symbols and formatting conventions used

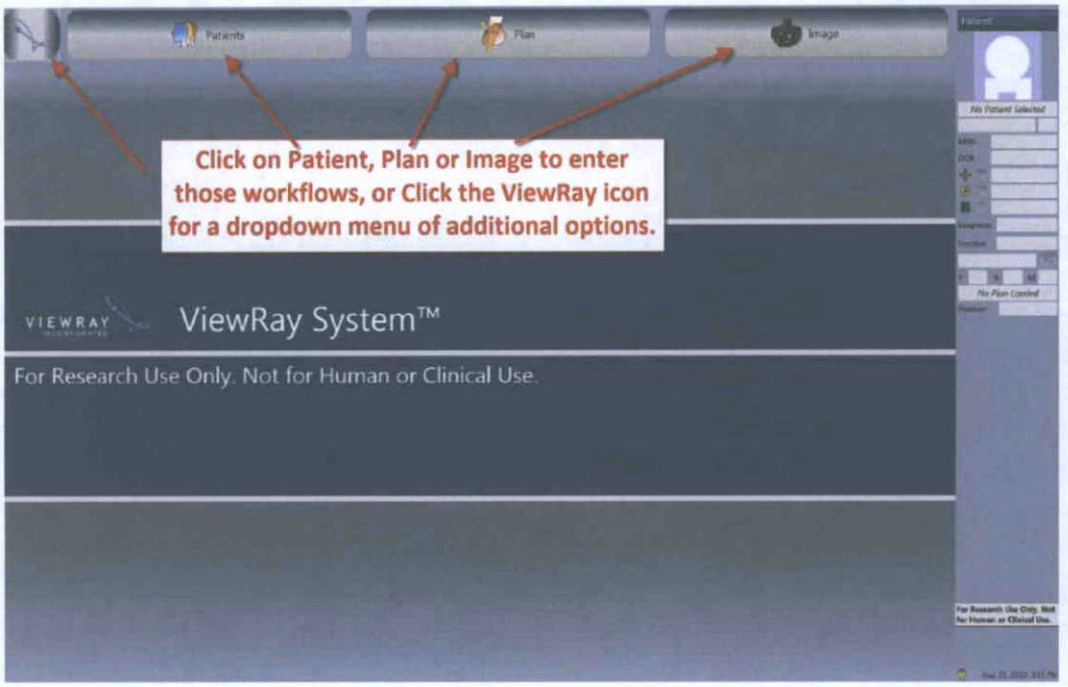
Numbered lists 1. 2. 3.	Task steps to be followed in sequence
❖	This symbol will appear before the description of an expected result.
Bold	Screens and button names, information boxes, menu items, and actions are all set in **bold**
❑ ➢	These two symbols introduce lists.
Bold italic	***Bold italic is used to emphasize important operating instructions***
>	An action sequence, for example: **Click > Plan**

THIS PAGE INTENTIONALLY LEFT BLANK

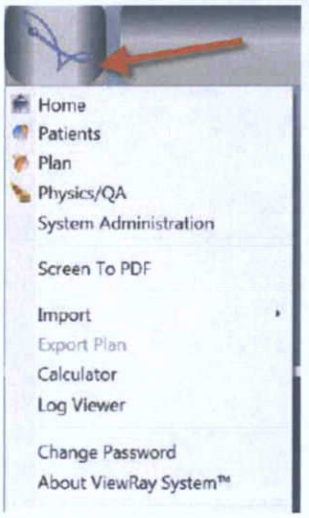
Chapter 1: Getting Started

1.1 Program Start Up

Treatment Planning is powered up from the front control panel power button on the CPU. To begin, push the control panel power button. The system will boot up and automatically start the treatment planning application. The Home screen will open.

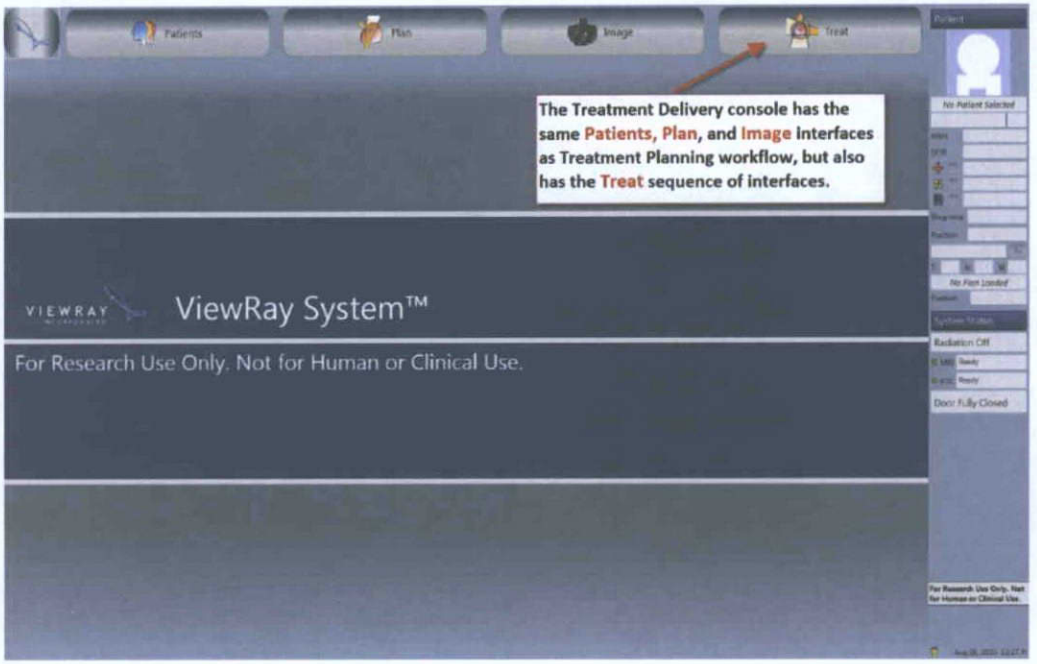


When you click the ViewRay™ icon a drop down menu opens. From this menu you can access additional interfaces and options, such as Physics/QA, System Administration, Import and Export plans, Log Viewer, Calculator, and Change Password.



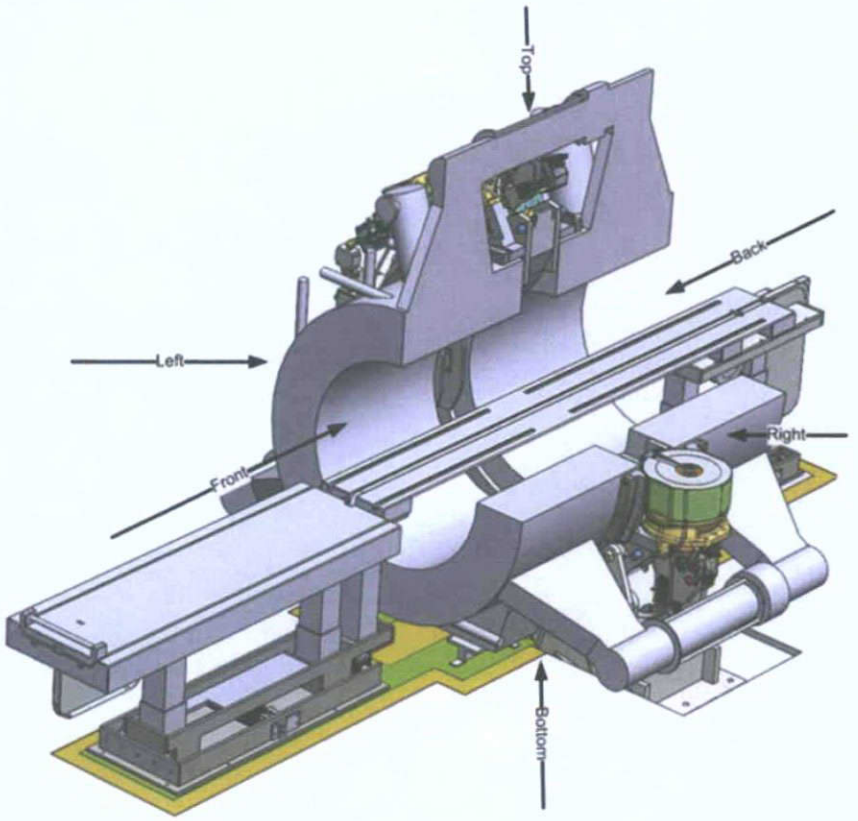
The **Treatment Delivery System** is also started from a power control panel. Press the on button on the power control panel. The system boots up and starts the application.

The **Treatment Delivery Home Screen** has **Patient, Plan, Image** and **Treat** options.



1.2 ViewRay System™ Coordinates

The ViewRay System uses 3-dimensional coordinate systems to fully specify the geometric location of the data throughout the various systems. This includes: cobalt sources, multi-leaf collimator positions, radiation beam trajectories, couch location and MR, CT, PET images. The coordinates' systems are described in Cartesian coordinates.



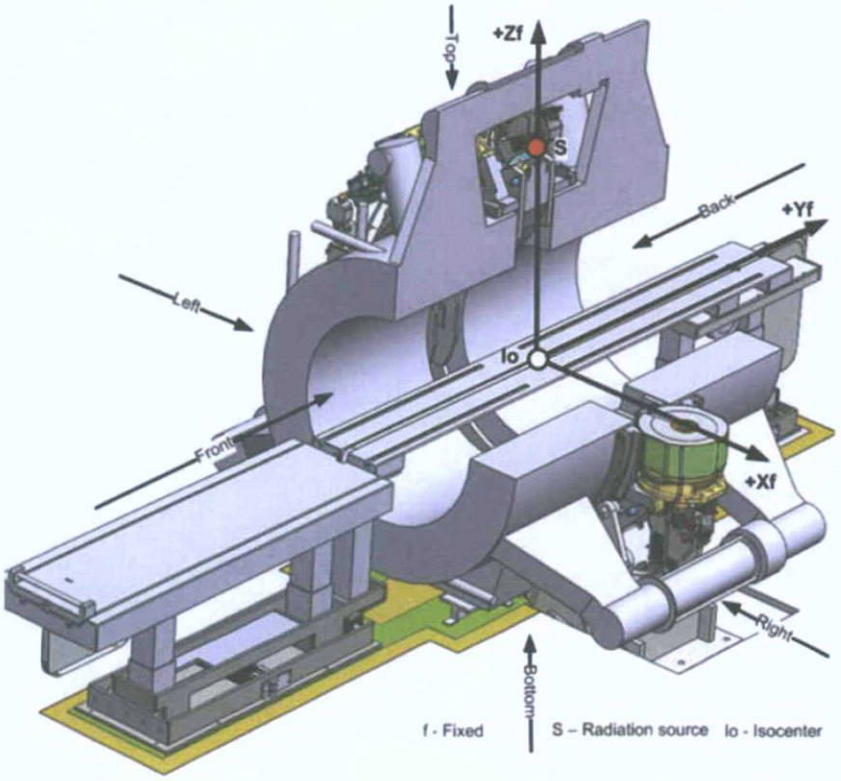
1.2.1 Direction Definitions

- **The front of the machine** is defined as the face of the machine where the couch enters the machine
- **The back of the machine** is defined as the face opposite the front of the machine
- **The top of the machine** is the face of the machine closest to the ceiling
- **The bottom of the machine** is the face of the machine closest to the floor
- **The left of the machine** is on the observer's left when the observer is standing facing the front of the machine
- **The right of the machine** is on the observer's right when the observer is standing facing the front of the machine

In accordance with IEC 61217, the ViewRay System has a fixed right-handed coordinate system with the coordinate system origin at the isocenter of the machine.

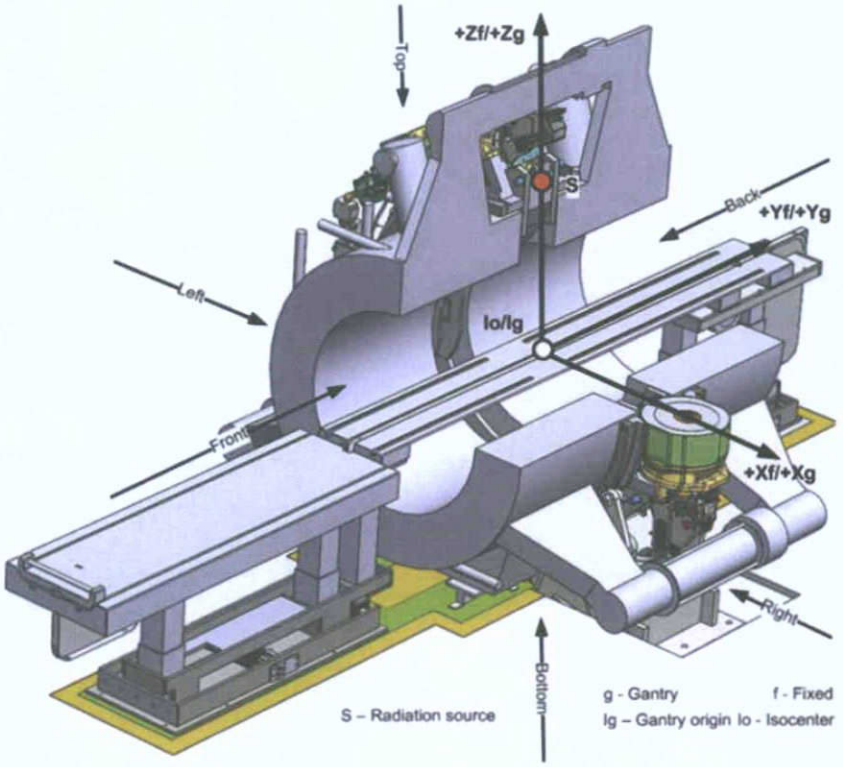
The "+Z" axis is aligned towards the top of the machine and the "-Z' axis towards the bottom of the machine. The "+Y" axis is aligned toward the back of the machine and "-Y" axis is aligned towards the front of the machine. The "+X" axis is aligned towards the right of the machine and the "-X" axis is aligned towards the left of the machine.

The radiation source location for the "0°" position is indicated by the red dot with label "S," that is the radiation source is aligned on the "+Z" axis. The isocenter is illustrated by the white dot with label "Io."



1.2.2 Gantry Coordinate System

The gantry at its "0°" position.

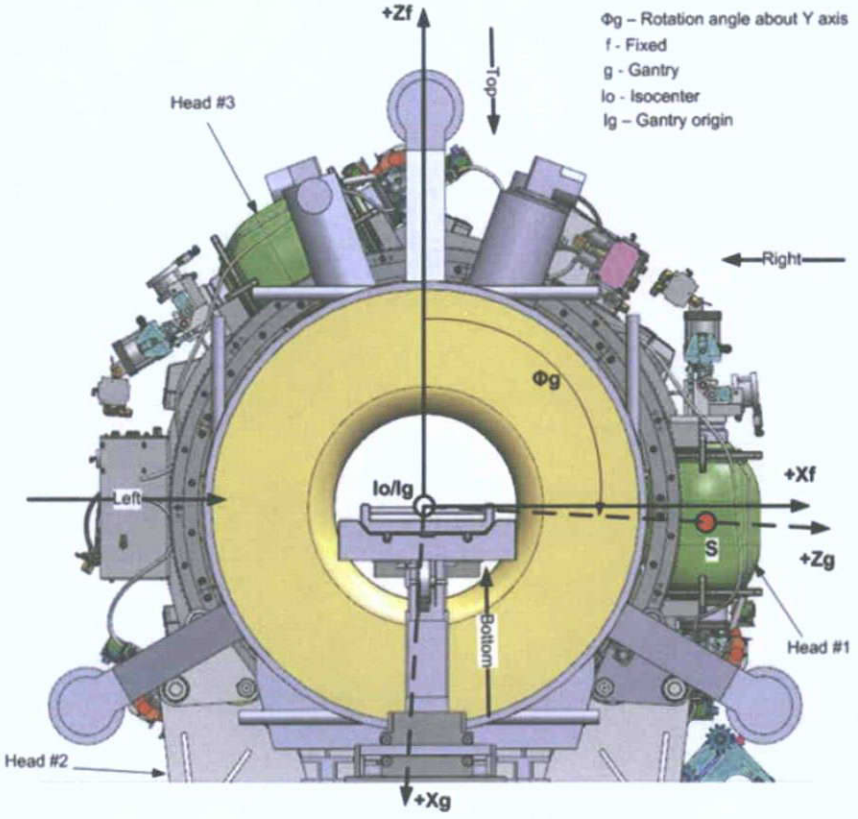


The gantry system is coincident with the fixed system and "Io" and "Ig" are at the same location. The fixed system is the mother system for the gantry system. The gantry coordinate system rotates with the gantry and can only rotate about the Y axis.

In this system the "+Z" axis is aligned towards the top of the machine and the "-Z" axis towards the bottom of the machine. The "+Y" axis is aligned toward the back of the machine and "-Y" axis is aligned towards the front of the machine. The "+X" axis is aligned towards the right of the machine and the "-X" axis is aligned towards the left of the machine.

1.2.3 Gantry Rotation

Positive (Clockwise) gantry rotation is illustrated below. The view is from the front of the gantry and the angle is slightly greater than 90°.



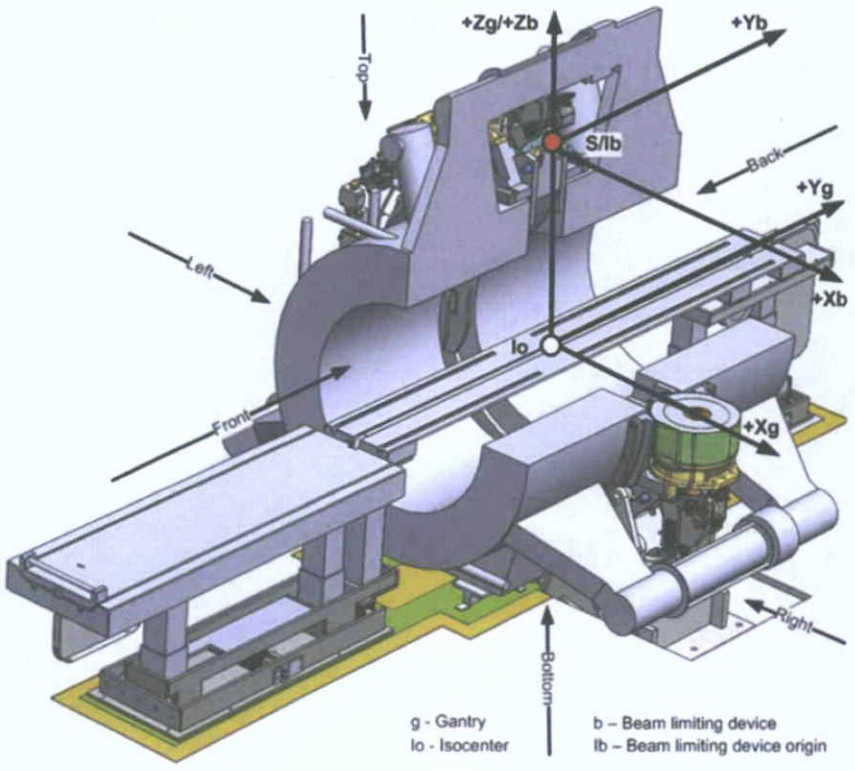
1.2.4 Beam Limiting Coordinate System

There are three beam limiting devices, also know as multi-leaf collimators, located 120 degrees apart from the other. The beam limiting coordinate system is the same as that of the gantry. The beam limiting device will rotate with the gantry. When one beam limiting device is at the "0°" position, its coordinate system is defined with the same axis orientation as the gantry system. The source "S" and origin "Ib" are at the same position.

The "+Z" axis is aligned to the top of the machine and the "-Z" axis to the bottom of the machine.

The "+Y" axis is aligned to the back of the machine and "-Y" axis is aligned to the front of the machine.

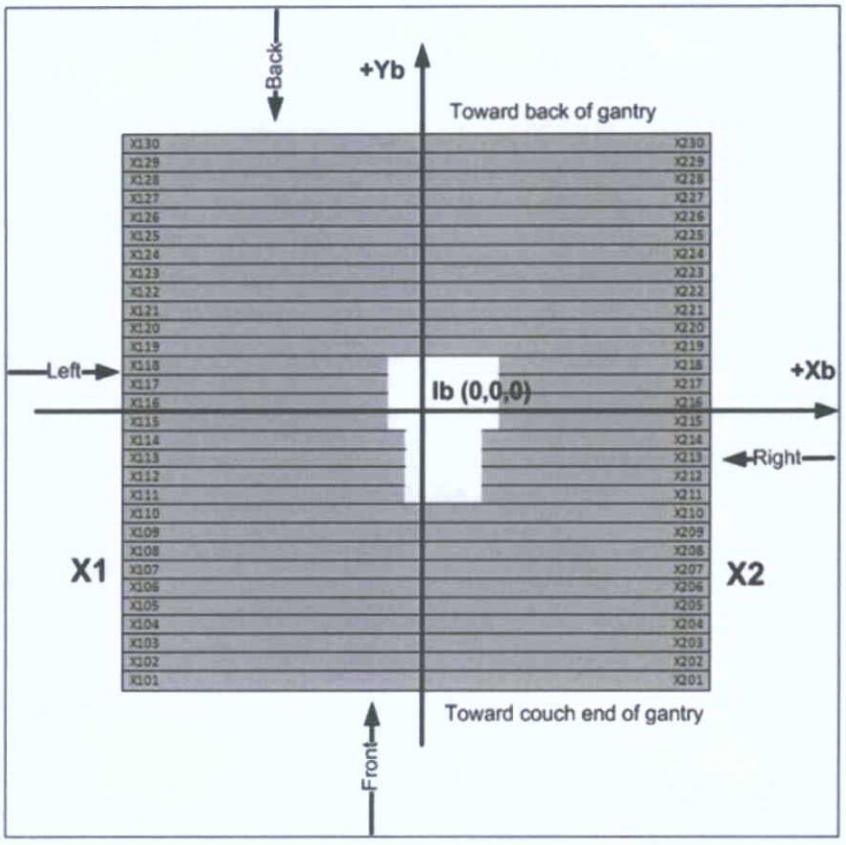
The "+X" axis is aligned to the right of the machine and the "-X" axis is aligned to the left of the machine.



1.2.5 Beam's Eye View of Multi-Leaf Collimator (MLC)

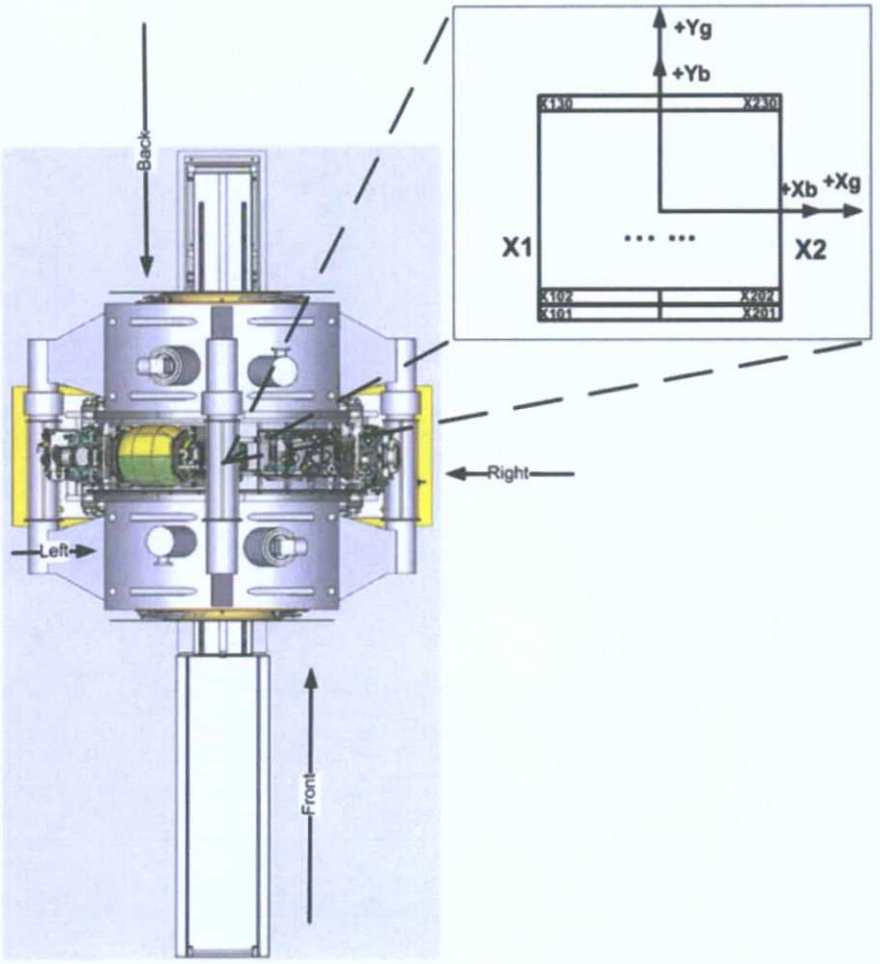
The movement of ViewRay's MLC is restricted to the plus and minus "X" directions.

The MLC leaves below show an open position of several of the leaves. The orientation of the beam limiting device is shown relative to the couch and gantry. This view of the MLC is considered a beam's eye view since we are looking toward the gantry isocenter.



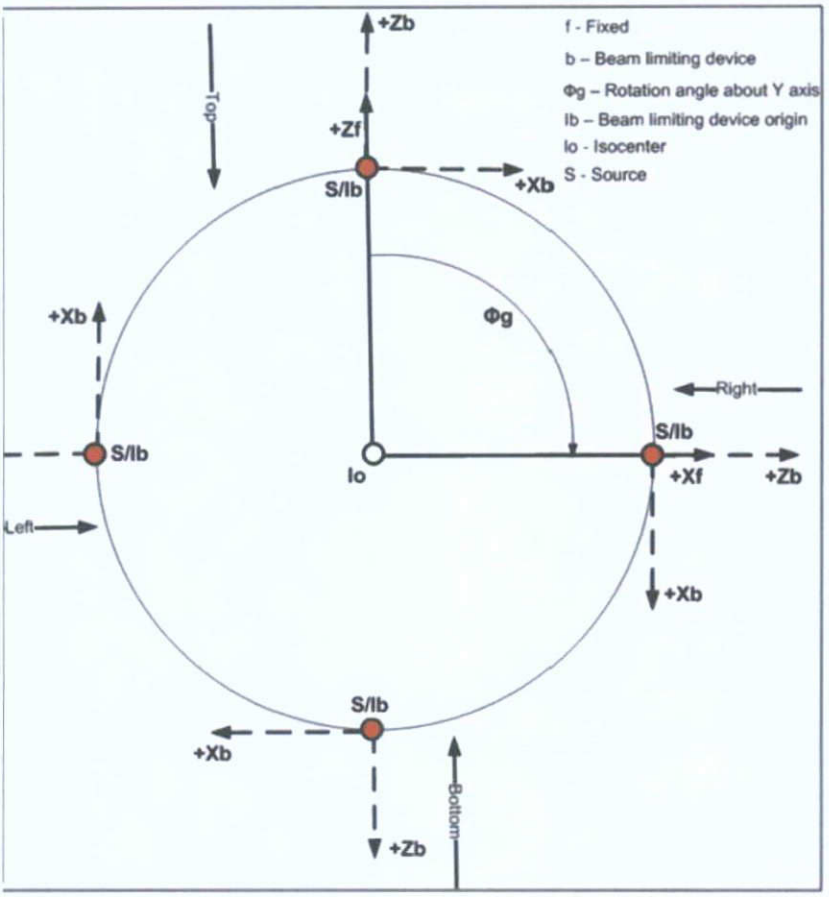
1.2.6 Orientation of an MLC Relative to the Gantry

The diagram below shows the orientation of an MLC relative to the gantry. The diagram is at the "0" gantry position and is looking down at the gantry from above.



1.2.7 MLC Orientation

The figure below shows the MLC and Beam limiting coordinate system for gantry angles of 0°, 90°, 180°, and 270°. *"ϕg" is only indicated for the 90°.*



1.2.8 Planning and Visualization using Coordinate System

The visualization system displays the Multi-planar reconstruction (MPR) planes as depicted in the three following diagrams.

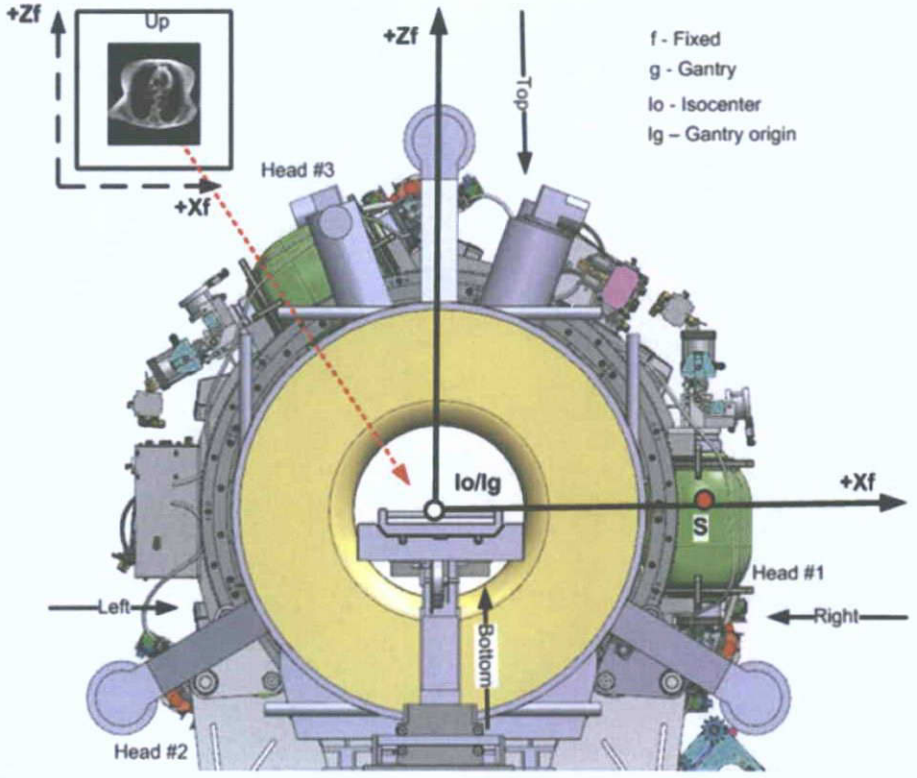
In each diagram the image is shown in a black square. The orientation on the user interface (UI) screen is the same orientation as the square in the top left corner of each diagram.

The axial, lateral and vertical images are always displayed in the fixed coordinate system. Note for all three orientations the patient is shown head first, supine, going into the gantry from the front.

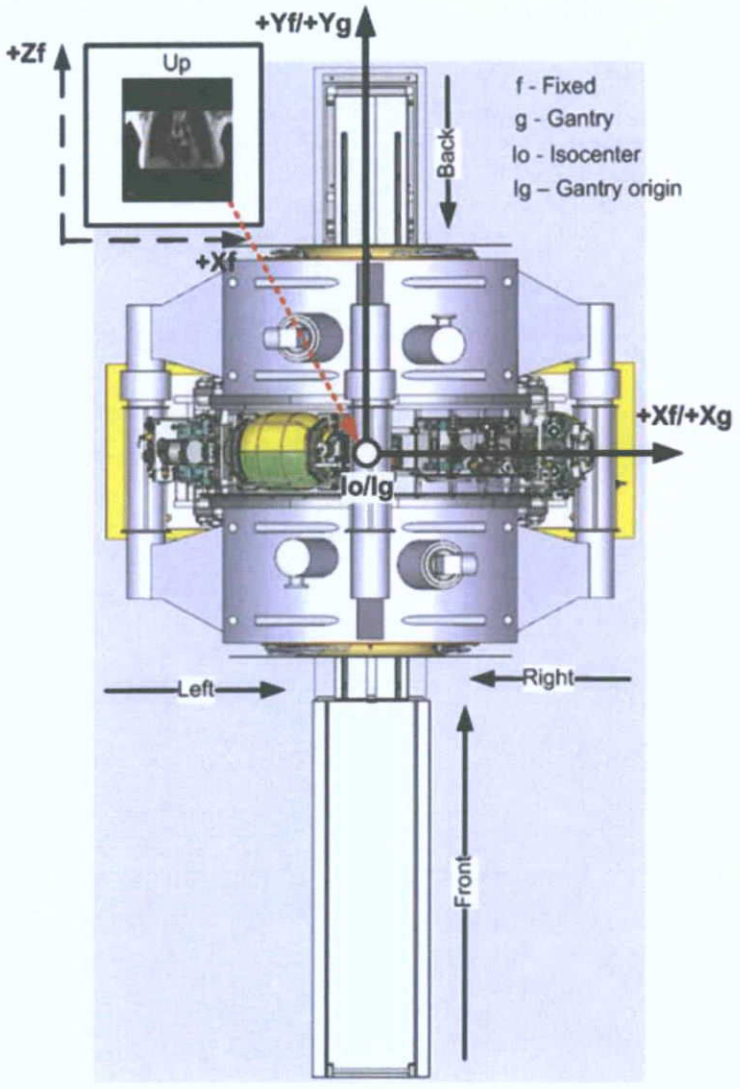
In this system the "+Z" axis is aligned to the top of the machine and the "-Z" axis to the bottom of the machine. The "+/-Y" axis is not visible. The "+X" axis is aligned to the right of the machine and the "-X" axis is aligned to the left of the machine.

The axial image is viewed as if looking into the gantry from the front.

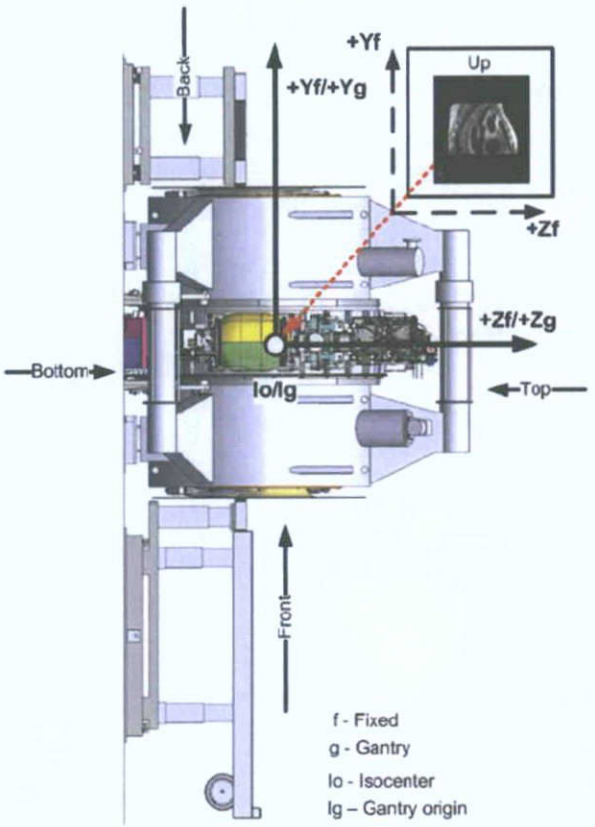
AXIAL MPR Image:



Vertical MPR Image:



Lateral MPR Image:

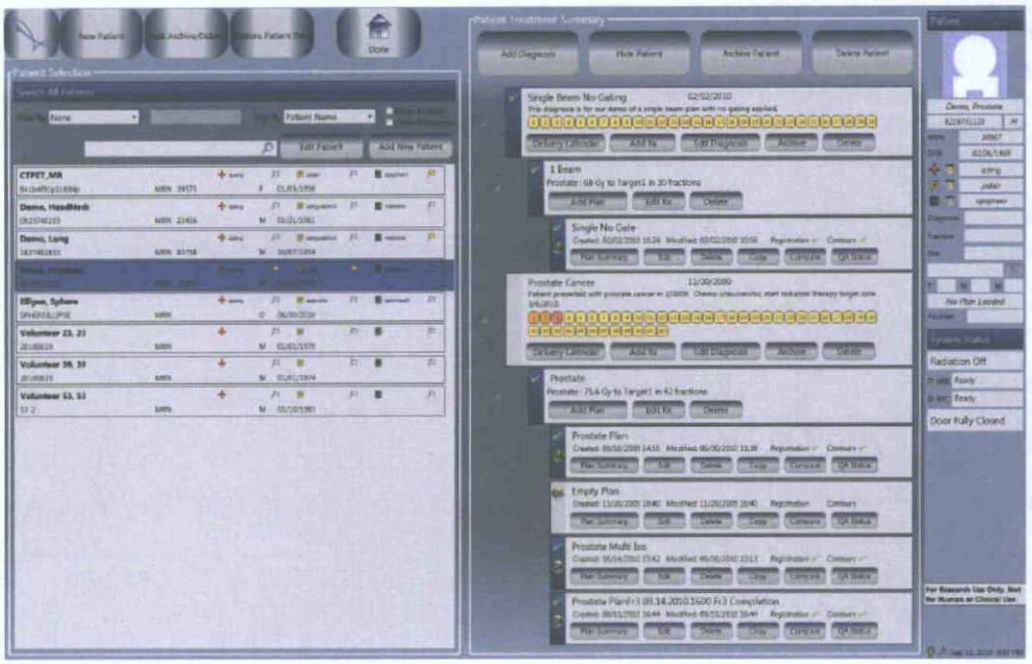


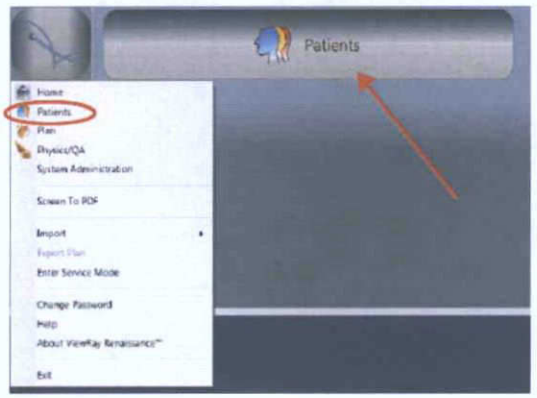
[redacted]

THIS PAGE INTENTIONALLY LEFT BLANK

Chapter 2: Prescription and Patient Management

The **Patients** interface is designed to record patient, prescription, and treatment data. Physicians can prescribe treatments, track doses, review treatment plans, or correct the course of patient treatments from the **Patients** interface.





You can access the **Patients** interface by using the dropdown menu on the **Home** screen or by clicking on the **Patients** button.

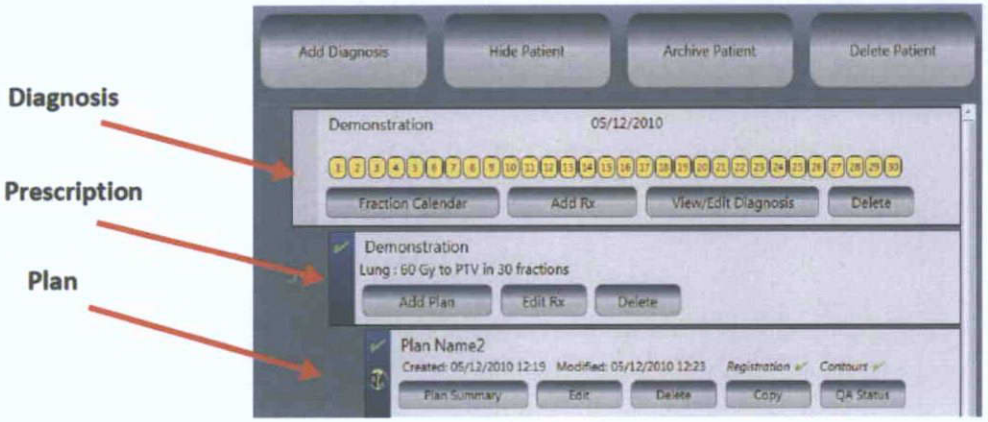
2.1 Patient Information Entry and Editing

2.1.1 Organization of Patient Management Information

Patient diagnosis, prescription, and plan are organized into three levels of selection boxes in the **Patient Treatment Summary.**

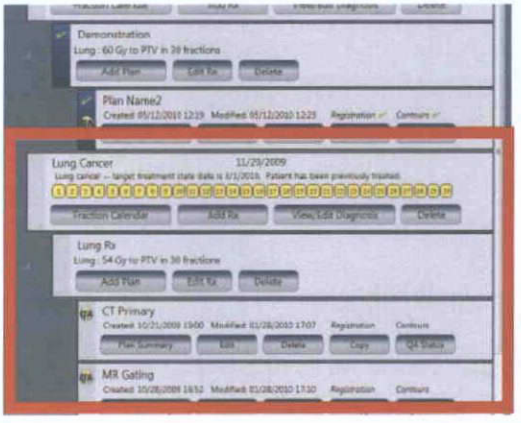
- **Diagnosis** is the first level selection box.
- **Prescription** is the second level selection box.
- **Plan** is the third level selection box.

❖ Multiple prescriptions and plans can be linked to a single diagnosis.



❖ Secondary diagnoses with their own associated prescriptions and plans can also be listed.

The second and third levels can be closed so that only the diagnosis (first) level is visible. To close a selection box Click the arrow at the left of the box.

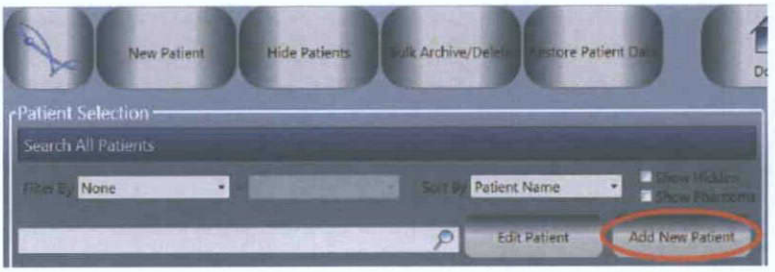


The View Ray System may be contraindicated for patients with implants due to the presence of the MRI magnetic field. All patients must be screened for MRI compatibility prior to treatment.

2.1.2 Adding New Patients

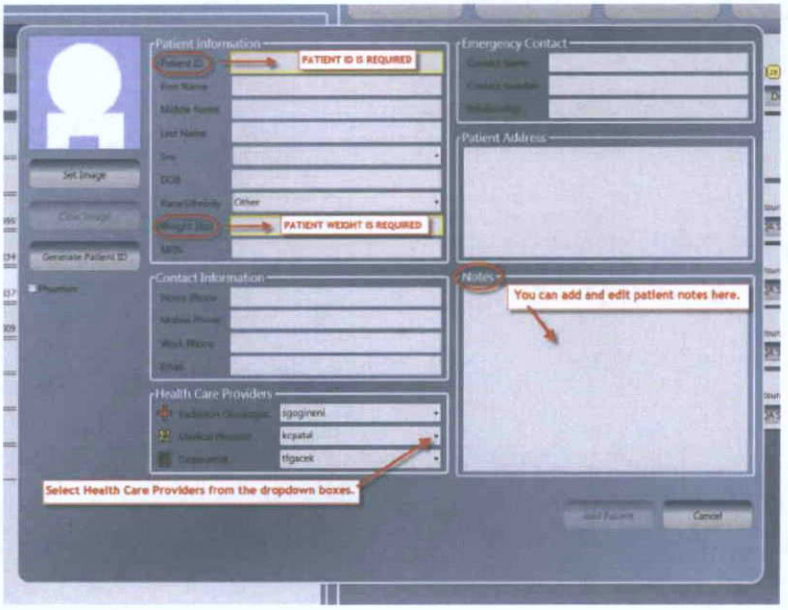
To add a new patient:

1. **Click > Add New Patient** from the Patient Selection group box.



❖ An information box will open.

2. Fill in the patient information.

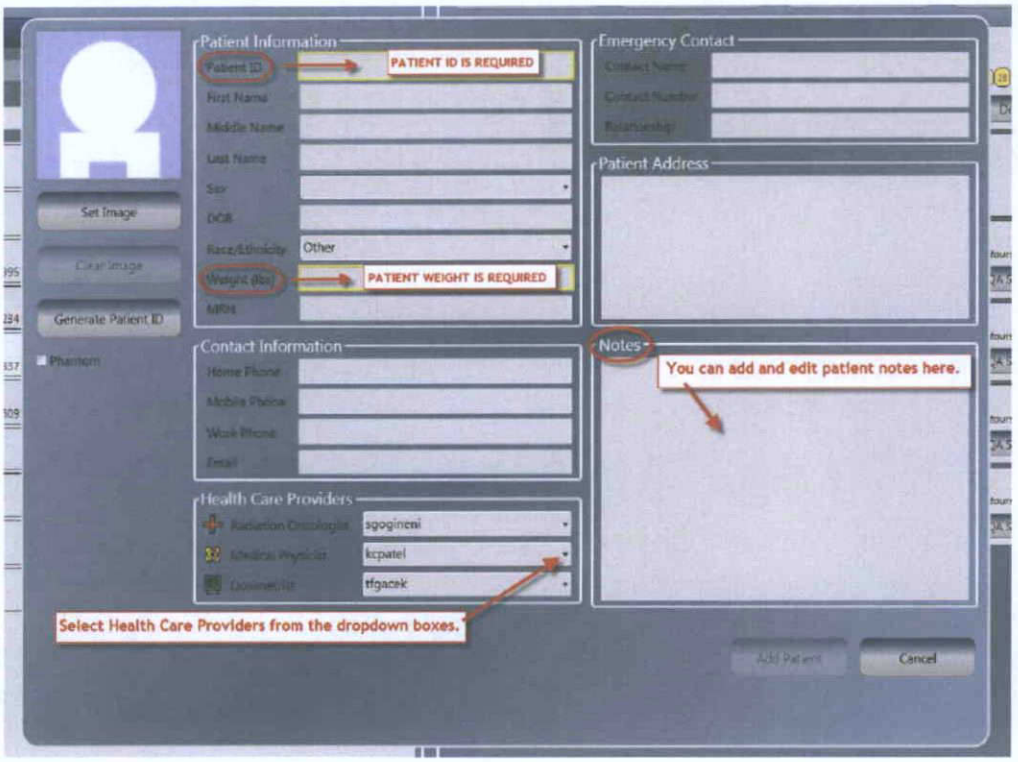


3. **Click > Add Patient** button on the bottom right of the information box.

2.1.3 Using the Patient Information Features

When setting up the patient information you can include designations for **Radiation Oncologist, Medical Physicist,** and **Dosimetrist** from dropdown boxes. When you do this you will be able to filter and sort the patients by any one of those categories.

You can also log notes and emergency contact information.



2.1.4 Filtering the Patient List

The patient list can be filtered by **Radiation Oncologist, Medical Physicist**, or **Dosimetrist.** Filtering will show only those patients who fit the filter criterion, that is, those who have either the same Physician or Physicist or Dosimetrist.

To use the **Filter By** feature:

1. **Click** > the **Dropdown Arrow** in the **Filter By** box
2. **Click** on the filter criterion you want to select.

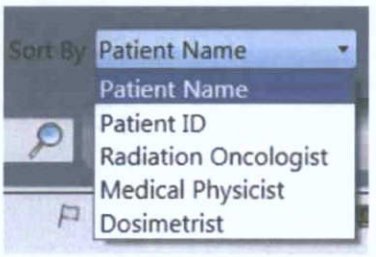
2.1.5 Sorting the Patient List

You can sort the patient list by **Patient Name, Patient ID, Radiation Oncologist, Medical Physicist,** and **Dosimetrist.**

The default sort is by **Patient Name.**

To change the sort:

1. **Click** > the **Dropdown Arrow** in the **Sort By** dropdown box.
2. **Double-click** on the **Sort By** criterion you want to select.



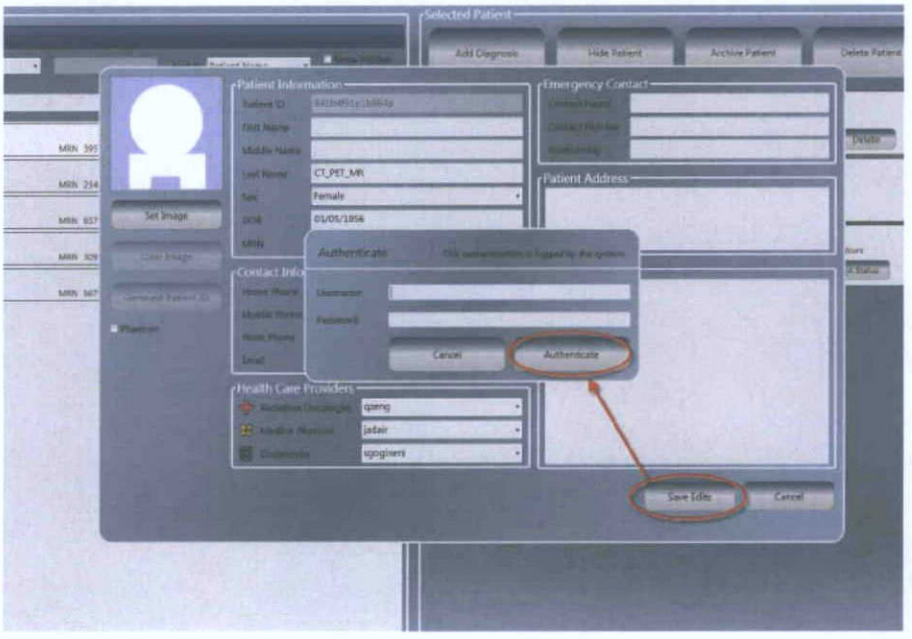
2.1.6 Editing Patient Information

To edit patient information:

1. Select a patient by Clicking on the patient from the list of patients.
2. Click > Edit Patient button.

- ❖ The Patient Information box will appear.

3. Add the edits to the patient information.
4. Click > Save Edits.
   - ❖ An authentication box will appear.
5. Enter your username and password in the authentication box.
6. Click > Authenticate.

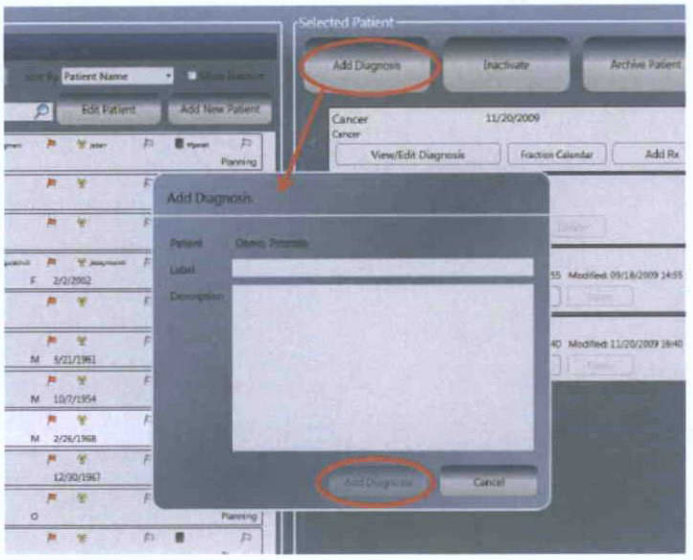


2.1.7 Adding a Diagnosis

To add a diagnosis to a patient's file:

1. Click > **Add Diagnosis** from the **Selected Patient** group box.

   ❖ An information box will appear.



2. Add the diagnosis Label and Description.
3. Click > **Add Diagnosis**.

   ❖ The diagnosis will be added to the **Selected Patient** list for that patient.

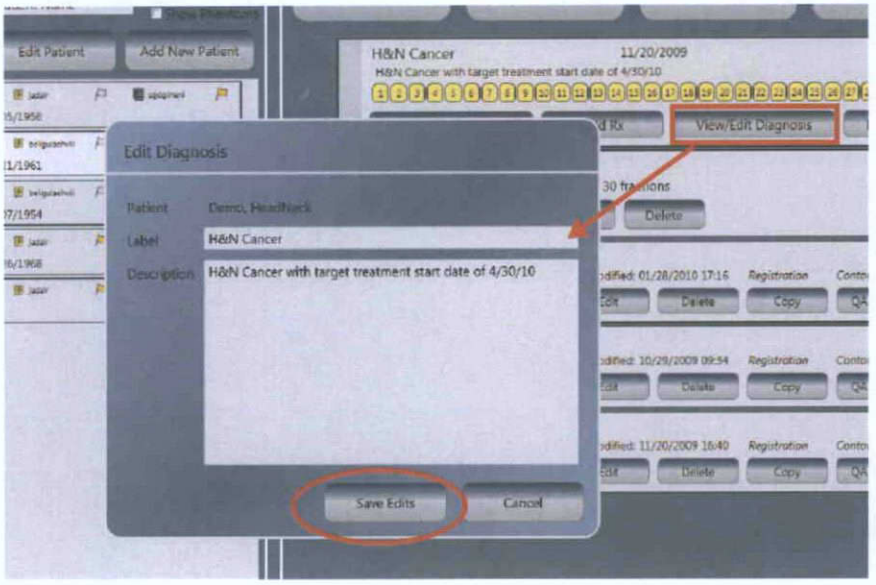
NOTE You must add a diagnosis before you can add a prescription. See Entering a Prescription on page 27.

2.1.8 Editing a Diagnosis

To edit a diagnosis:

1. **Click > View/Edit Diagnosis** from the **Diagnosis** selection box.

   ❖ An **Edit Diagnosis** information box will open in the center of the screen.



2. Add your edits.

3. **Click > Save Edits.**

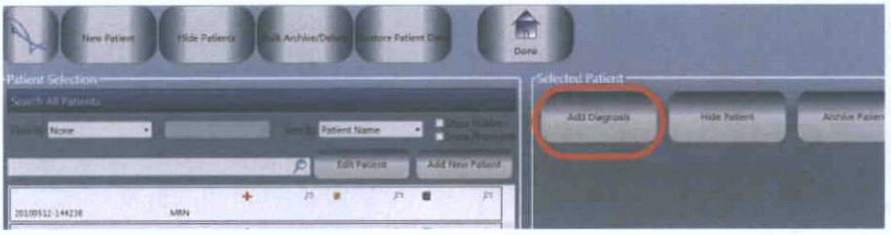
2.1.9 Entering a Prescription

A stub prescription which requires a patient ID, label, and disease site must be filled out to begin treatment planning. The remaining fields on the prescription form are not required until prescription approval. See Adding a Prescription on page 59.

You can enter a prescription for a new patient or add a prescription for a patient already in the system.

For instructions on entering a new patient See See Adding New Patients on page 21.

A newly added patient will appear at the top of the list of patients in the **Patient Selection** group box.

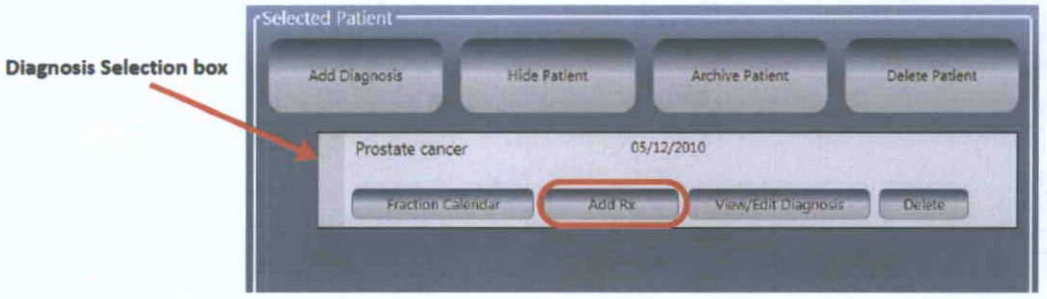


NOTE Before you can add a prescription for a new patient you must have or add a diagnosis. See Adding a Diagnosis on page 25.

- A newly added diagnosis will appear at the top of the list in the **Selected Patient** group box which is on the right half of your screen.

To add a prescription to the diagnosis:

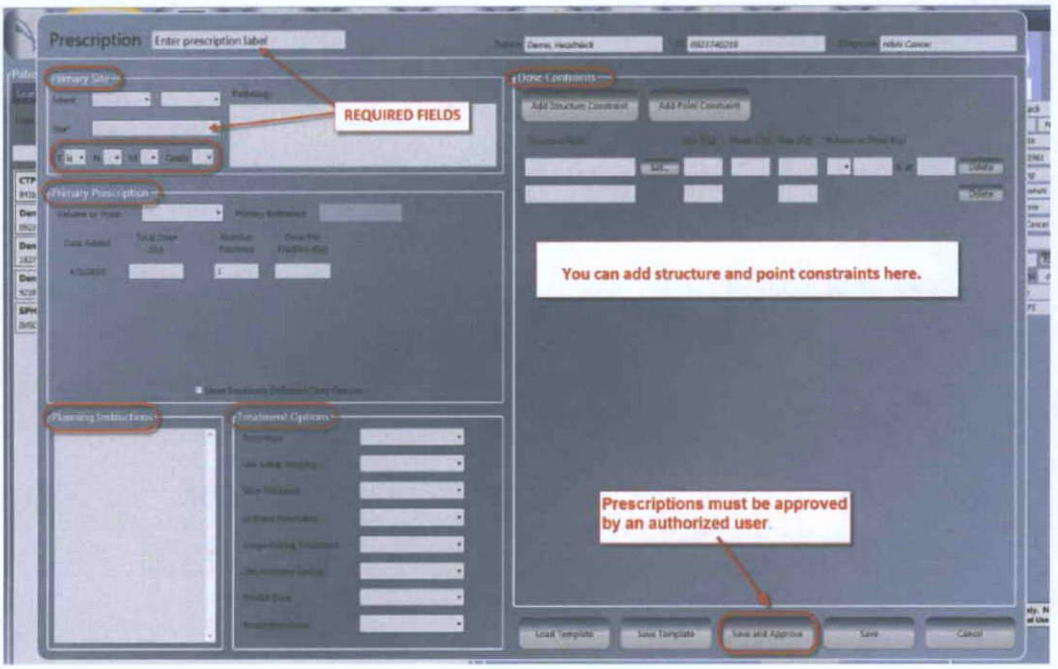
1. **Click > Add Rx** from the Diagnosis selection box.



- The prescription information box will open.

Prescription label and site are required fields in the Prescription information box.

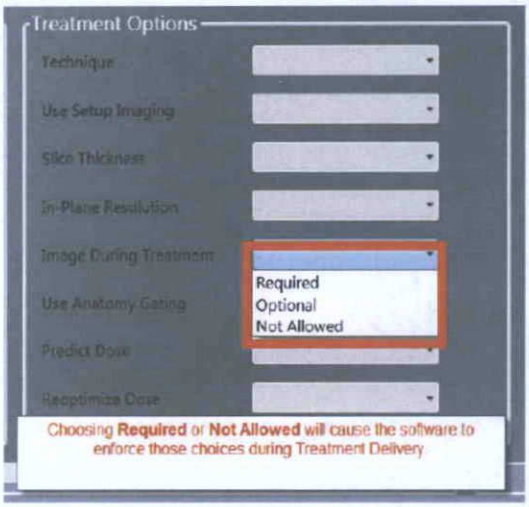
In addition to the required fields, you can enter a primary prescription, add dose, point and structure constraints, Tumor, Node and Metastasis (TNM) staging, planning instructions, and treatment options. See Adding Structure, Dose, and Point Constraints on page 61. Also see See Entering TNM Values on page 68.



2.1.10 Using Treatment Options

The treatment specifications you choose in the **Treatment Options** dropdown boxes are enforced during treatment delivery.

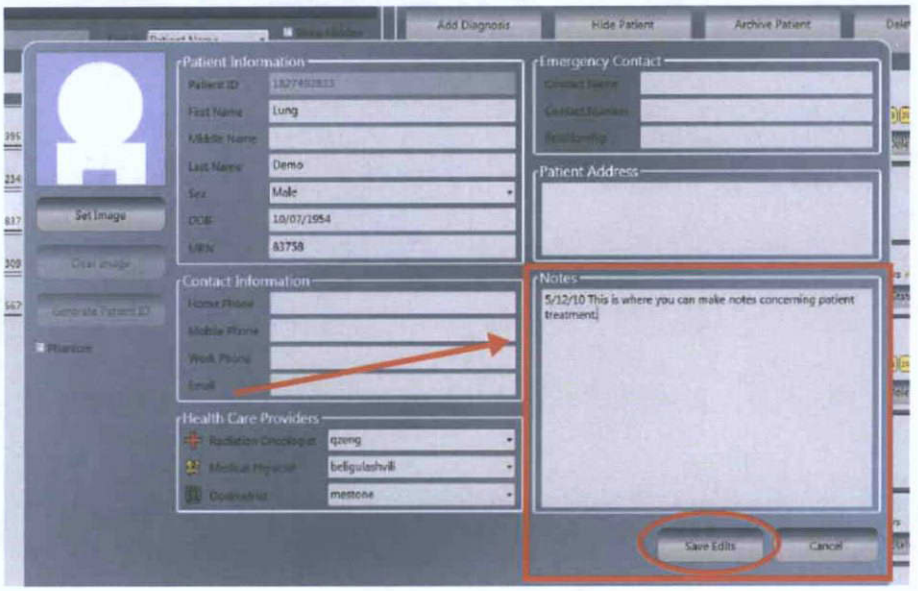
If an option is marked as **Required** you will not be able to proceed in the workflow interface until that treatment option is enabled.



2.1.11 Editing Patient Notes

1. **Click > Edit Patient**

❖ **The Patient information box will open**



2. **Add your notes to the Notes section of the information box.**
3. **Click > Save Edits**

❖ **An authenticate prompt appears.**

4. **Enter username and password**
5. **Click > Authenticate**

2.1.12 Archiving Patients

Inactive patients can be removed from the active list of patients and archived to a folder where they can easily be retrieved if necessary.

You can archive individual patients or you can select several patients and perform a bulk archive.

To individually archive a patient:

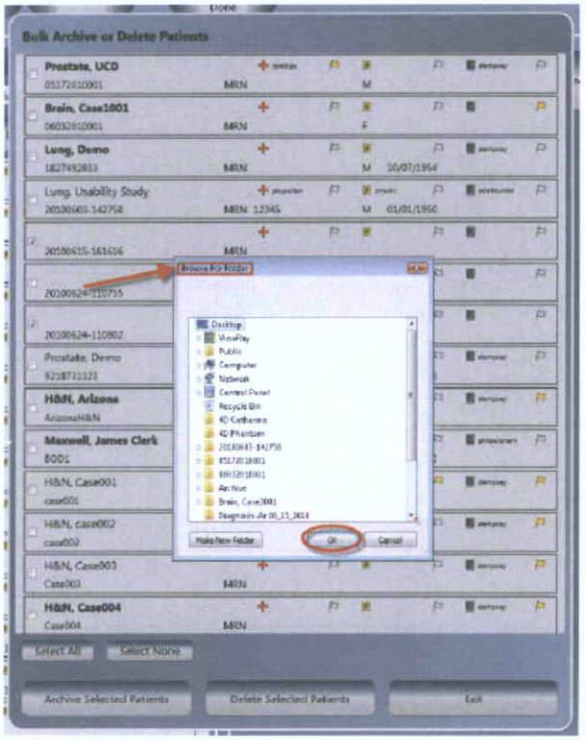
1. **Click > Archive Patient** button at the top of the **Selected Patient** group box.

   ❖ **The Browse for Folder** information box will open.

2. Select the destination folder you want to use to archive the patients.
3. **Click > OK**

To use the bulk archive:

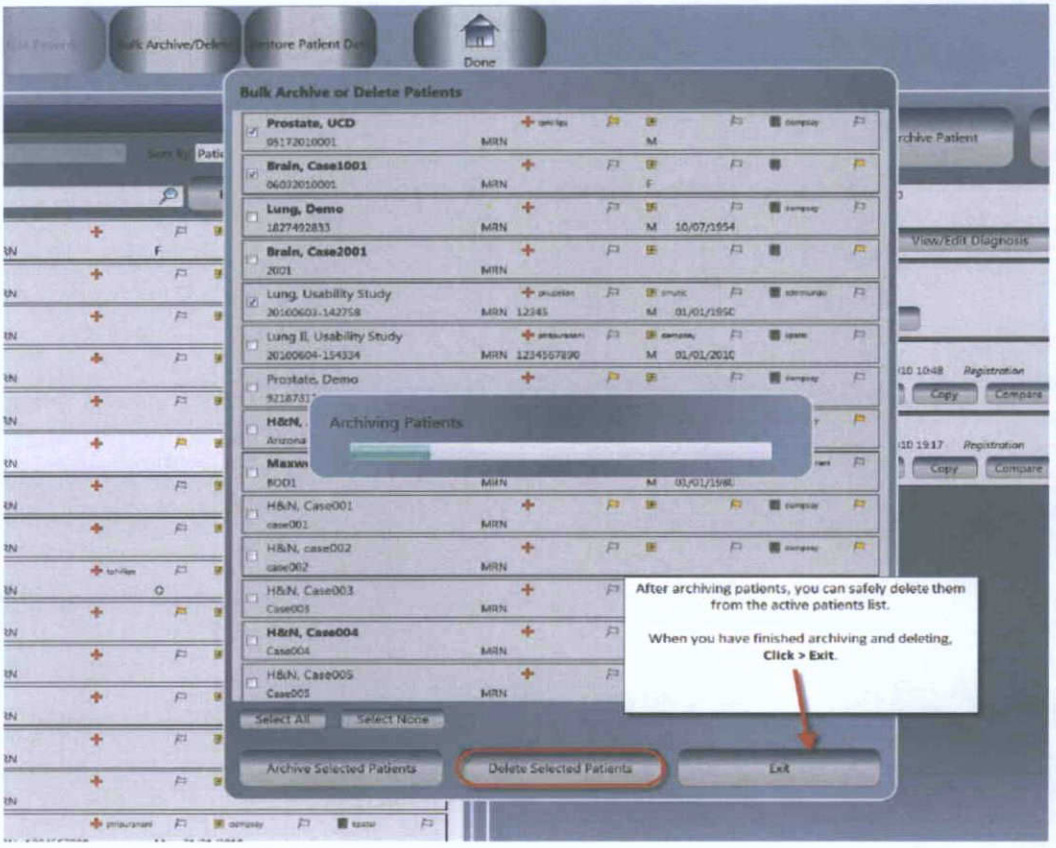
1. Click > **Bulk Archive/Delete** button at the top of the screen.
   - ❖ The **Bulk Archive or Delete Patients** window will open in the center of the screen.
2. Check the box next to any patients you want to archive.
3. Click > **Archive Selected Patients** button at the bottom of the **Bulk Archive or Delete Patients** window.
   - ❖ The **Browse for Folder** information box will open.



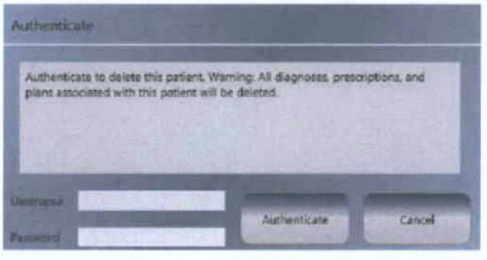
4. Select the destination folder you want to use to archive the patients.
5. Click > **OK**

❖ A progress bar will appear in the center of the screen.

4. When the progress bar disappears, you can exit the screen by **Clicking > Exit** or you can safely delete the archived patients from the active list by **Clicking > Delete Selected Patients.**



You must authenticate, with your username and password, the deletion of any patients.



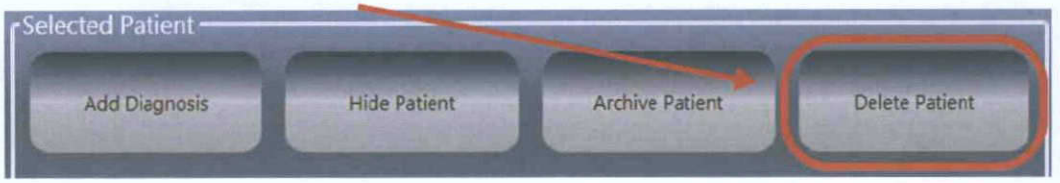
2.1.13 Deleting Patients

Before you delete patients make sure you have archived them so you can retrieve those patients and their plan data in the event that is necessary.

There are two ways to delete patients. You can delete a single patient or you can choose to delete several patients at once using the **Bulk Archive/Delete** feature. See Archiving Patients on page 31.

To delete an individual patient:

1. **Select a patient by Clicking** on the patient name from the **Patient Selection** list.
2. **Click > Delete Patient**



- An authentication information box will open in the center of the screen with a warning that deleting the patient will delete all associated plans and images.

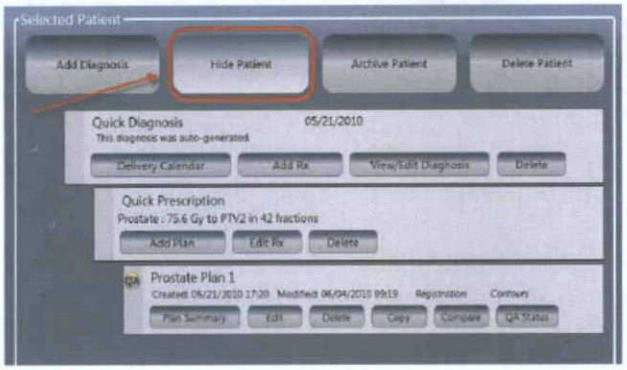
3. To authenticate the deletion, enter your username and password
4. **Click > Authenticate**

2.1.14 Hiding Patients

To hide a patient from the active list:

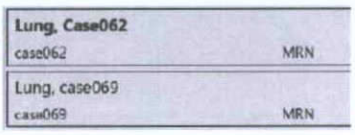
1. **Click** on the patient's name from the **Patient Selection** group box.
   - ❖ The patient's diagnosis and any associated prescriptions and plans will appear in the **Selected Patient** group box.

2. **Click** > **Hide Patient** button from the **Selected Patient** group box.



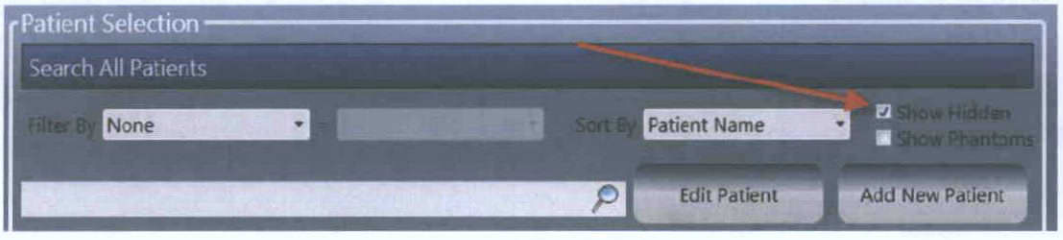
   - ❖ The listing for the patient will change from a boldface font to a regular weight font and the **Hide** button will change to **Un-Hide**. The patient will not be hidden from the list until you close out from the interface and re-enter.

Hidden Patient will appear unbolded



To view hidden patients:

1. **Click** > **Show Hidden** checkbox (on top right of the **Patient Selection** group box).

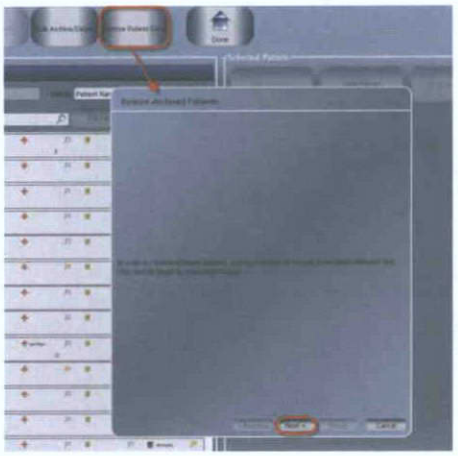


2.1.15 Restoring Patient Data

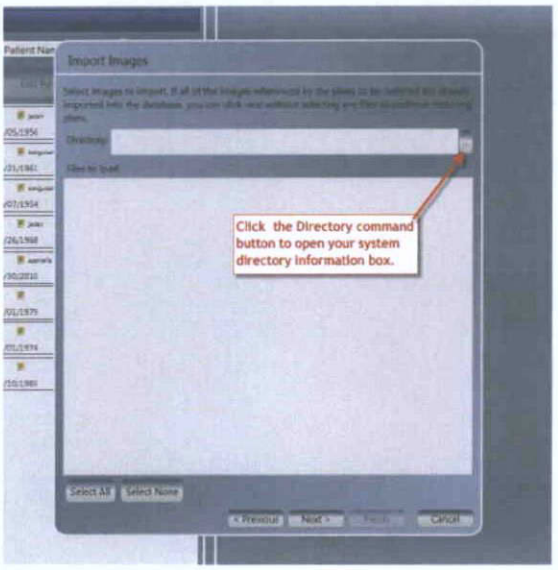
You can restore patient data that has been archived.

To restore archived patient data:

1. Click > **Restore Patient Data** command button at the top of the **Patients** interface.
   - ❖ The **Restore Archived Patients** box will open.
2. Click > **Next.**



❖ The Import Images box will open.



3. Click > **Directory command button** (side button with dots).
   - ❖ Your system directory information box will open.

4. Choose a directory.

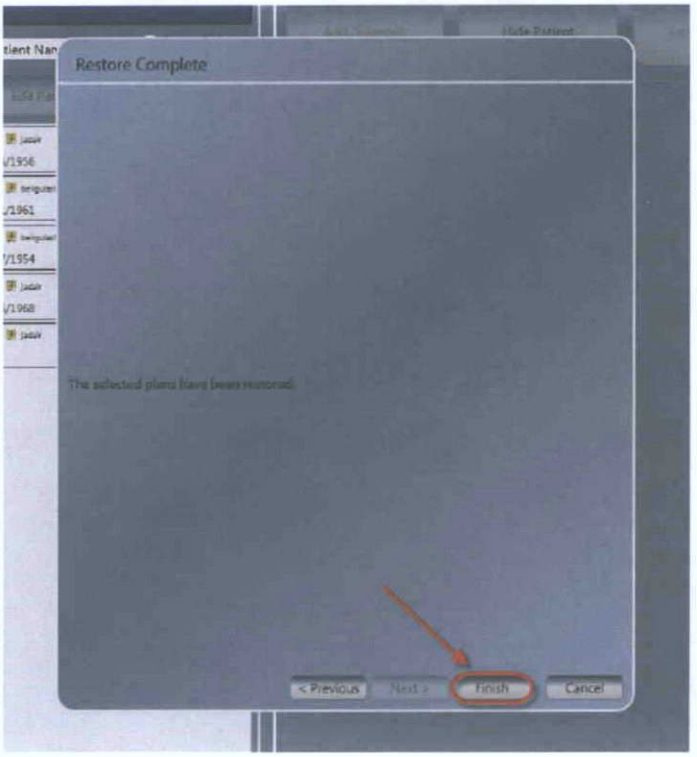
   ❖ The files from that directory will populate the **Files to load** section.

5. Click the file(s) containing the patient data you want to restore.

4. **Click > Next.**

   ❖ The **Restore Complete** confirmation box will open.

5. **Click > Finish.**



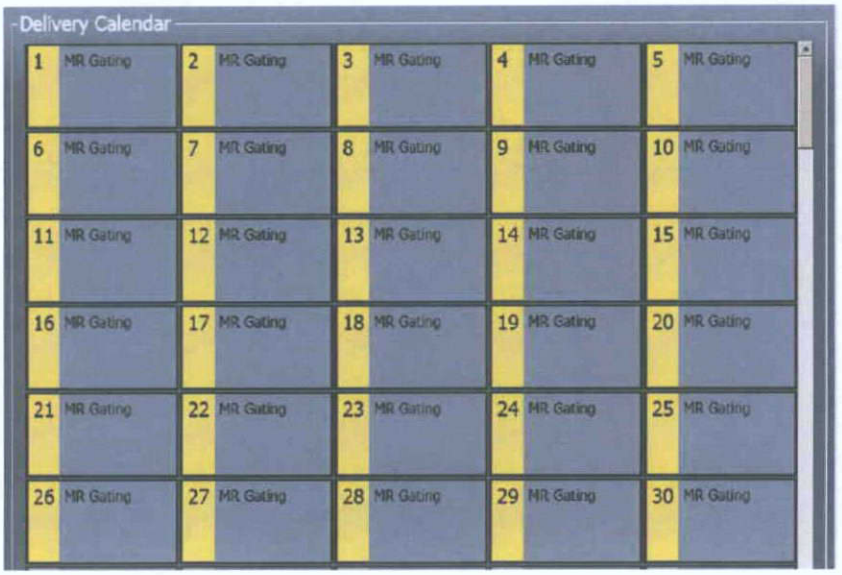
2.2 Scheduling Treatment

All treatment delivery fractions are scheduled using the Delivery Calendar.

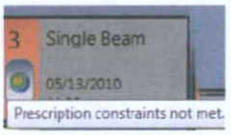
2.2.1 Viewing the Delivery Calendar

The Delivery Calendar stores and retrieves the treatment planning data and settings that will be used to deliver approved radiotherapy treatments.

The calendar also records and stores the history of each patient's treatment. When fractions are delivered, the plan name, date and time of the fraction, and the dose delivered are recorded in the calendar.



Undelivered fractions have a yellow line going through the fraction number.



Delivered fractions all have a **Beam Icon** beneath the fraction number. Delivered fractions that ended early or did not meet the prescription constraints have a pink line going through the fraction number.

Successfully delivered fractions have an orange line through the fraction number and **Beam Icon**.

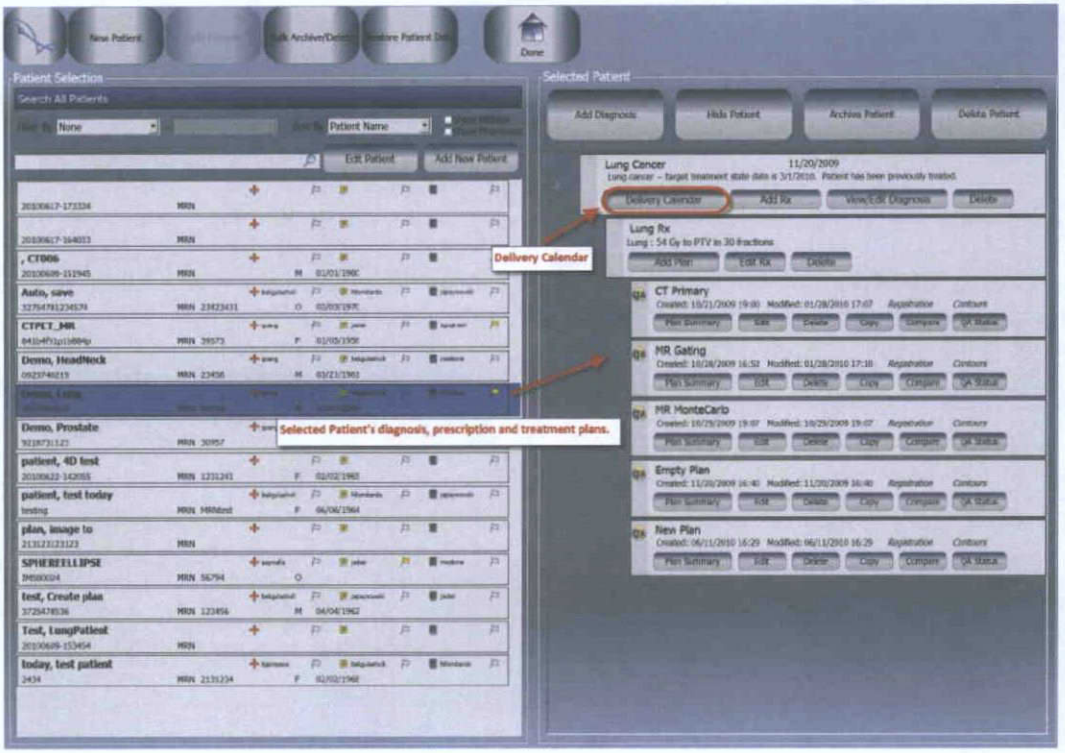
2.2.2 Assigning Treatment Plans to Fractions

Only approved treatment plans can be added to a fraction calendar. You can access both approved and unapproved plans from the **Patients** interface.

To assign a treatment plan to fractions:

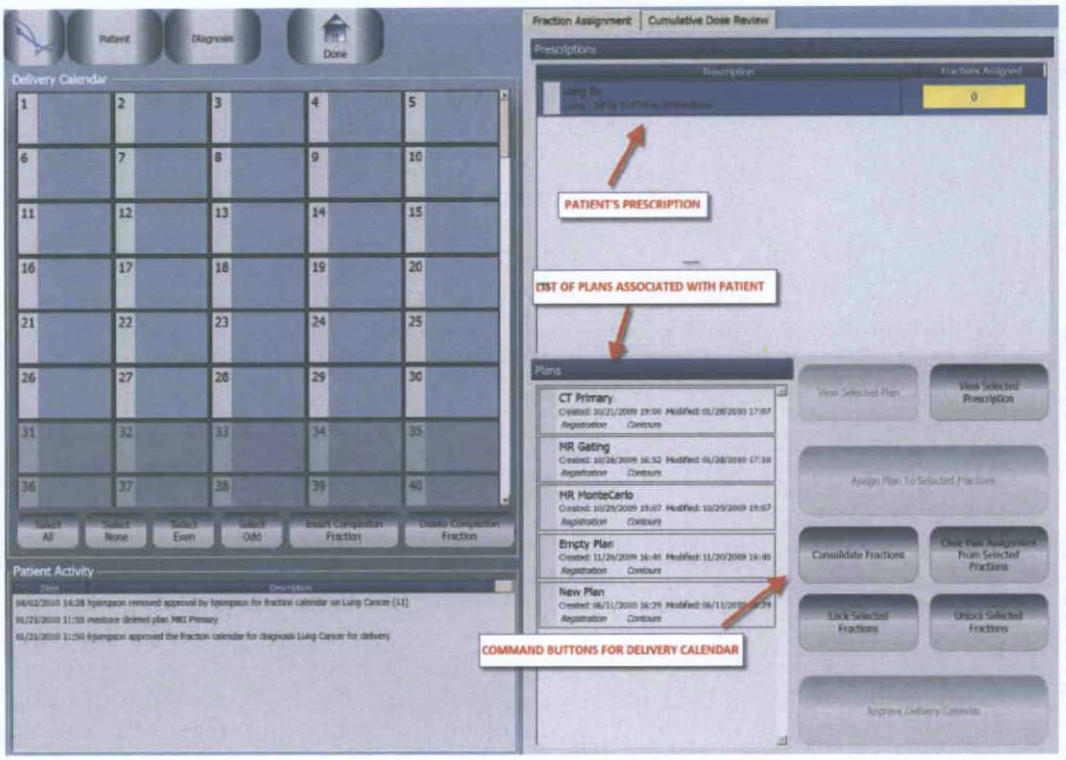
1. From the **Patients** home screen select a patient.

   ❖ The selected Patient's diagnosis, prescription, and plans will appear under the **Selected Patient** group box.



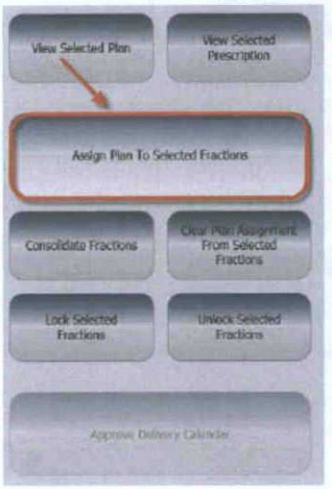
2. Click > **Delivery Calendar**.

   ❖ The **Delivery Calendar** opens.



3. Select the prescription.
   - ❖ Plans for that prescription will load and appear in the **Plans** list.
4. Click on the plan you want to assign to fractions.
5. **Hold down > Shift key and drag the mouse to Highlight** the number of fractions on the **Delivery Calendar** you want to **select.**
6. **Click > Assign Plan to Selected Fractions** from the command buttons on the right side of the screen.

- The plans are assigned to the **Delivery Calendar** and listed by Plan name as in the calendar example shown below.



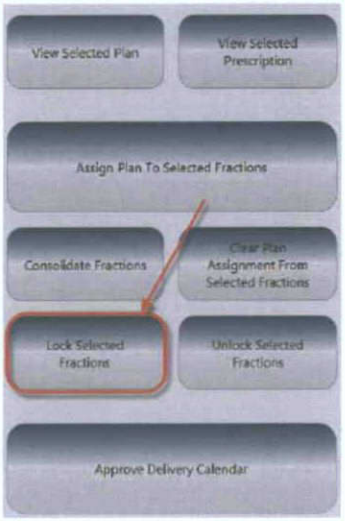
2.2.3 Locking Treatment Fractions

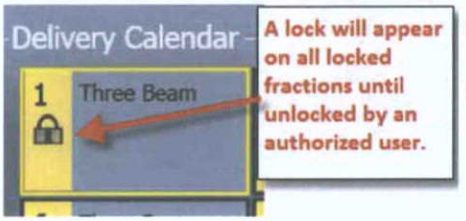
You can lock fractions so they cannot be delivered unless unlocked by an authorized user.

1. **Select the fraction or fractions you want to lock by Clicking on one fraction or select several fractions by holding down the Shift key and clicking on one fraction and then drag to select others.**

2. **Click > Lock Selected Fractions.**

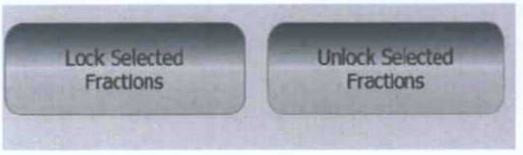
   ❖ **An authentication information box will open and you must enter your username and password to proceed with locking the fraction(s).**





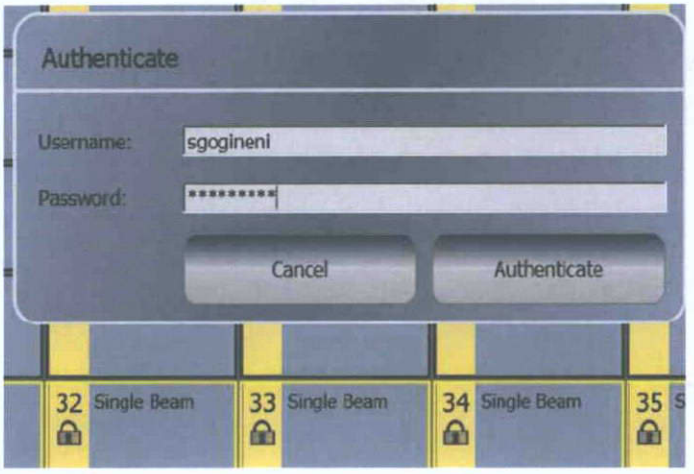
2.2.4 Unlocking Treatment Fractions

To Unlock Treatment Fractions:



1. From the Delivery Calendar, Select the fraction(s) you want to unlock.
2. Click > Unlock Selected Fractions.

❖ An **Authenticate** information box opens.



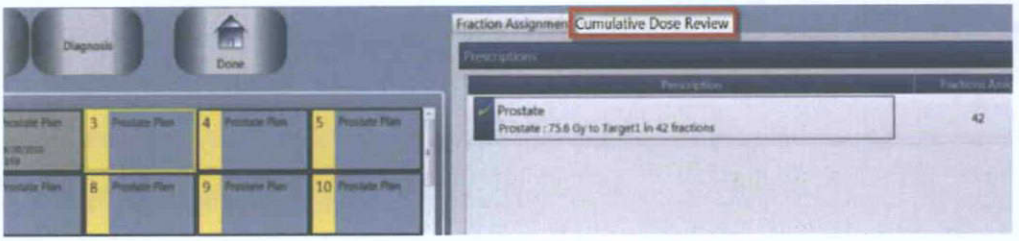
3. Enter your username and password.

❖ The lock symbols are removed from the fractions and they are enabled for treatment delivery.

2.3 Viewing Cumulative Dose Review

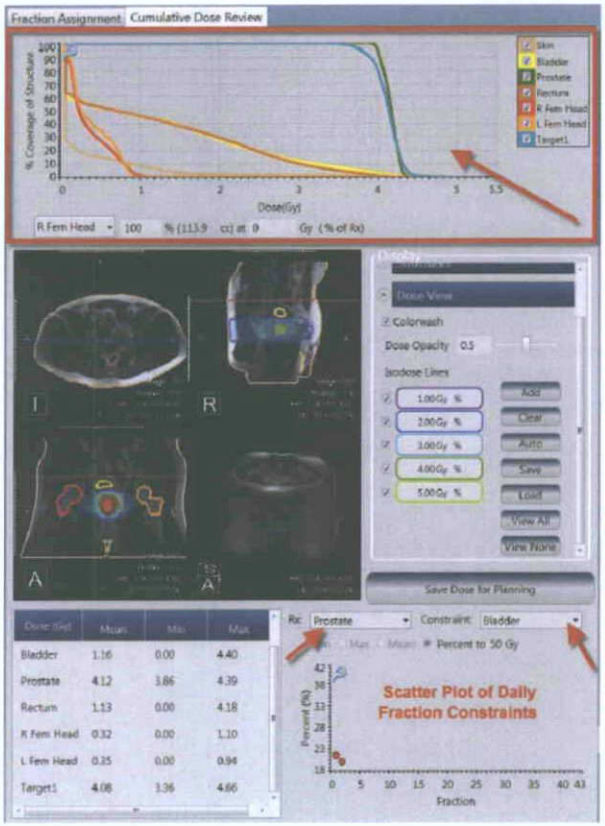
Any time during any patient's course of treatment you can view the patient's **Cumulative Delivered Dose.**

1. Click > Cumulative Dose Review.



❖ A progress bar will open.

When the progress bar closes, the cumulative DVH and additional dose measurement and visualization tools will be available.



2.4 Reviewing Individual Fraction Deliveries

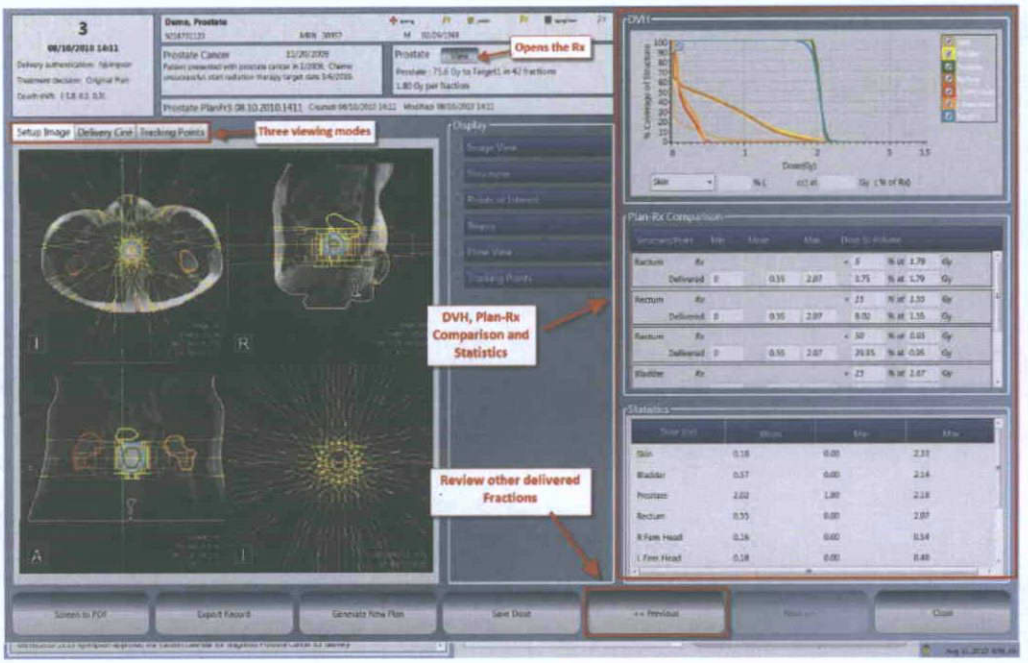
After a fraction has been delivered you can view the delivery review of that individual fraction:

1. **Click > Delivery Calendar.**

❖ The Delivery Calendar will open.

2. **Click > the Beam Icon (It appears on all delivered fractions).**

❖ A **Loading Plan** progress bar will appear. When the progress bar closes the **Fraction Review** UI will open.

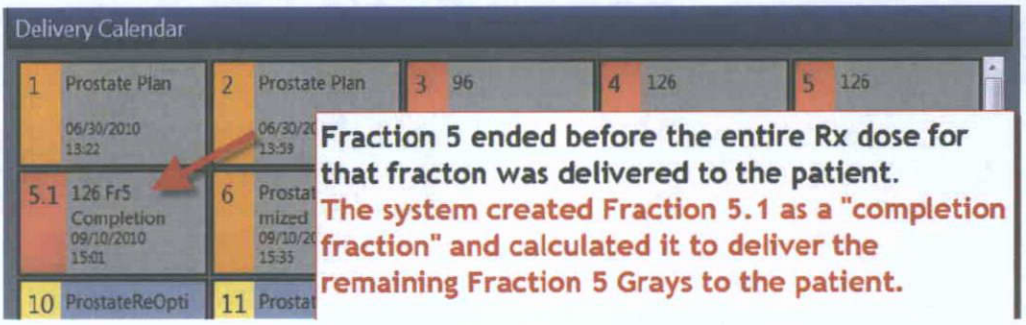


The fraction review has the **DVH, Plan-Rx Comparison**, and **Statistics** on the right side of the screen. By **Clicking the View** button on the screen you can view the prescription. There are three tabs above the images--**Setup Image, Delivery Ciné,** and **Tracking Points.**

- ❑ **Setup Image** is displayed on the fraction review UI when it opens.
- ❑ **Click > Delivery Ciné** to view a treatment delivery playback.
- ❑ **Click > Tracking Points** to view a chart of amplitude and periodic frequency. See Viewing the Tracking Points Chart on page 48. And, See Tracking Points on page 147.

2.5 Completion Fractions

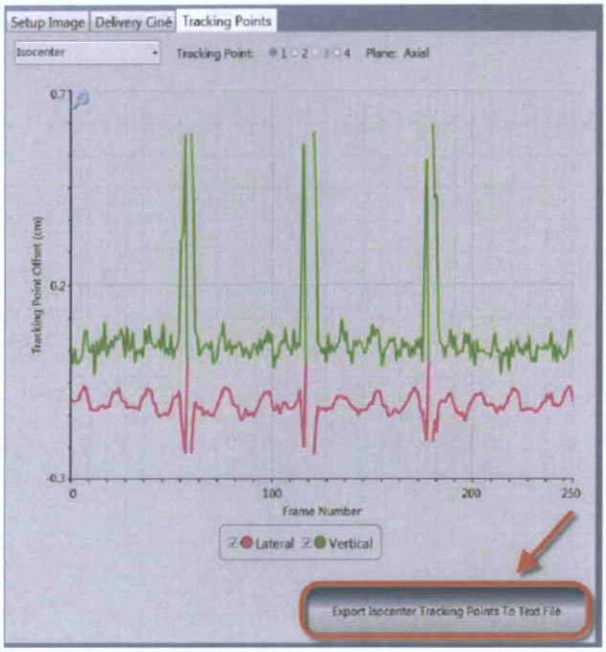
If a fraction ended early for any reason, the system will create a completion fraction calculated to deliver the remaining Grays. The completion fraction will be numbered with the number of the fraction not completed along with a decimal separator and the number 1.



2.6 Viewing the Tracking Points Chart

From the Fraction Review interface **Click > Tracking Points.**

❖ **The Tracking Point chart will open.**



The amplitude of the wave is the distance the point moved in a given frame. The horizontal axis represents time. Each frame comprises 0.25 second. You can zoom in on the chart by clicking and dragging the mouse. You can export the tracking point data by clicking the button in the lower right corner.

2.7 Setting QA Status

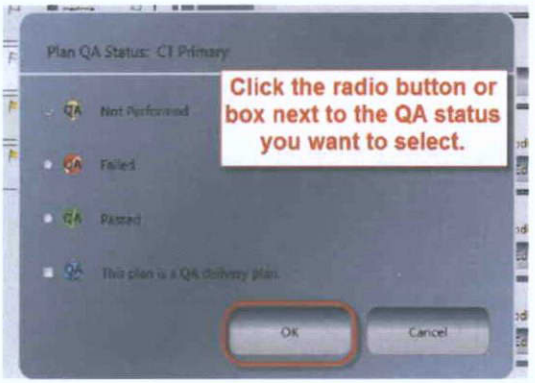
There are three QA status settings for plans and a **QA Delivery Plan** status:

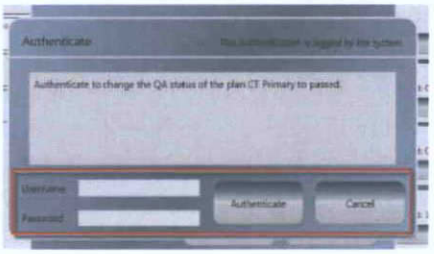
- Not Performed
- Failed
- Passed
- This plan is a QA delivery plan

The QA status of each plan is visible from the **Selected Patient** group box.

To change the QA status:

1. **Click > QA Status.**
   - ❖ The QA status selection box will open.
2. Check the option button or box you want the QA status changed to.
3. **Click > OK.**





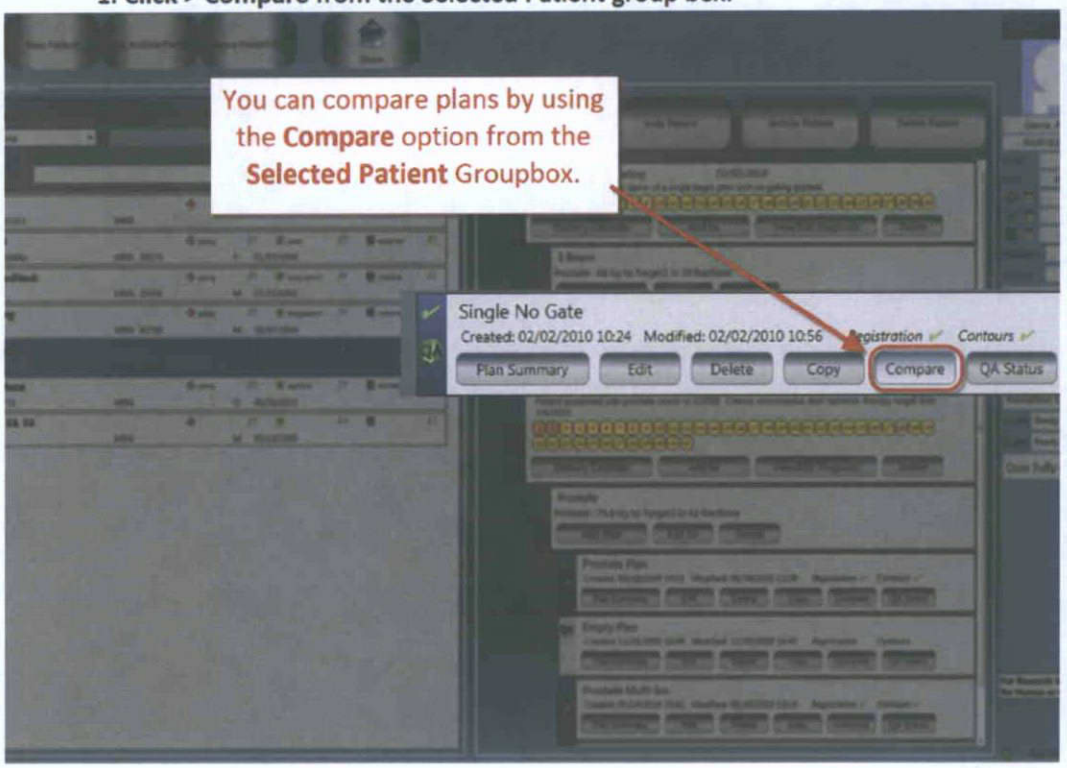
- ❖ An authentication information box will open.

4. Enter your username and password.
5. **Click > Authenticate.**

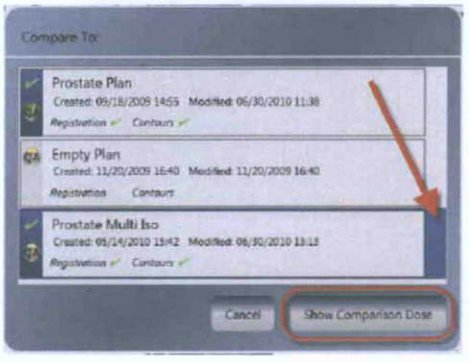
2.8 Comparing Plans

To compare plans:

1. Click > Compare from the Selected Patient group box.



❖ The Compare To information box will open.

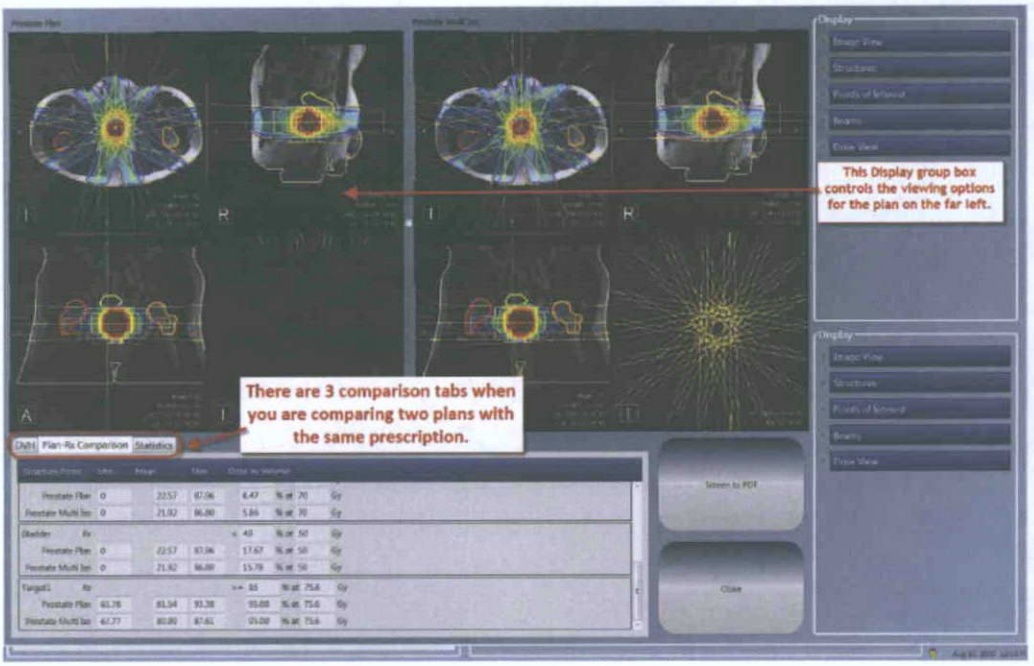


2. Click on the plan you want to compare your current plan to.

   ❖ The selected plan will be highlighted as indicated by the arrow.

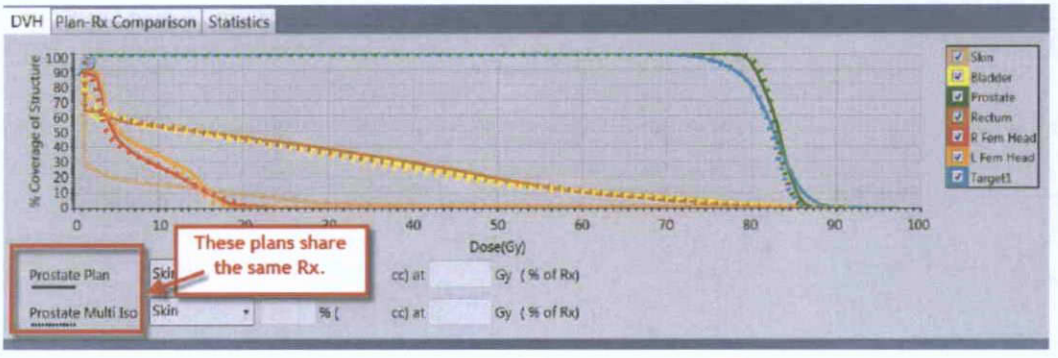
3. Click > Show Comparison.

The two plans you are comparing will appear side-by-side. There are two **Display** group boxes on the right side of the screen. The top Display group box is for the plan on the far left. The bottom Display group box controls the views for the next plan.



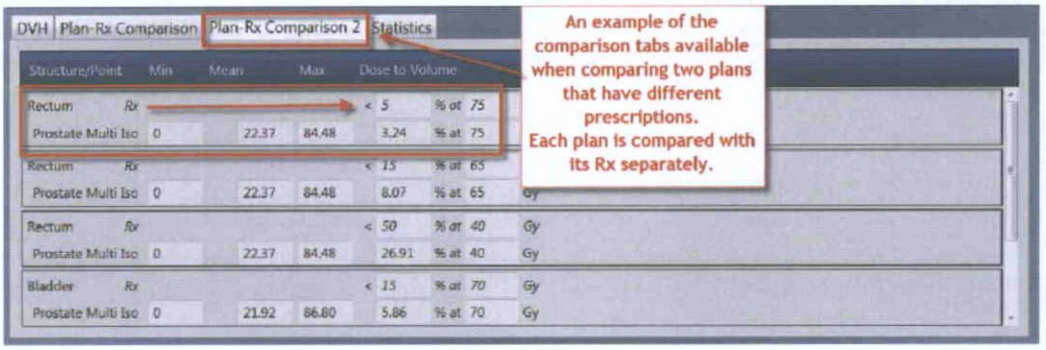
When comparing plans that have the same Rx, there are three comparison tabs you can **Click** on to view:

- **DVH**
- **Plan-Rx Comparison**
- **Statistics**



When you compare plans that have different prescriptions, there are four comparison tabs.

- **DVH**
- **Plan-Rx Comparison**
- **Plan-Rx Comparison 2**
- **Statistics**



Chapter 3: Treatment Planning

3.1 Treatment Planning Introduction

Delivering radiation to a patient requires that the radiotherapy dosimetrist and physician define the target volume, determine beam directions and shapes, calculate the associated dose distribution, and evaluate that dose distribution. The ViewRay System™ assists the radiotherapy planner and physician in this planning process.

The ViewRay System™ radiotherapy treatment planning system consists of a software program and a computing hardware platform. The treatment planning process includes a wide spectrum of tasks—from an evaluation of the need for imaging studies and continuing to an analysis of the accuracy of daily fractional treatments.

3.1.1 The ViewRay System Treatment Planning Process

The following tasks are available as part of the planning process:

1. **Image Acquisition and Input**

- Acquire and input CT, MR, and PET imaging information into the planning system.

2. **Anatomy Definition**

- Define and display contours and surfaces for target and critical structures.
- Geometrically register all input data (CT, MR, PET), including registration with initial simulation contours, films, patient position, etc.
- Define target contours, generate a 3-D target surface using surface expansion, import target information from multiple imaging modalities.
- Generate electron density representations from CT data or from assigned bulk density information.

3. **Beam/Source Technique**

- Determine beam arrangements.
- Generate beam's-eye-view displays.
- Design field shape (MLC).

4. **Dose Calculations**

- Select dose calculation algorithm and methodology
- Perform dose calculations.
- Set relative and absolute dose normalizations.
- Input the dose prescription.

5. **Plan Evaluation**

- Generate 2-D and 3-D dose displays.
- Perform visual comparisons.
- Use DVH analysis.
- Use automated optimization tools.

6. **Plan Implementation**

- Generate hardcopy output.
- Transfer plan to treatment machine.

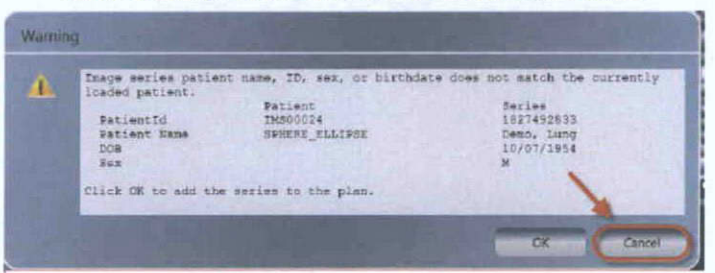
7. **Plan Review**

- Perform overall review of all aspects of the plan before implementation.

The following sections of the Treatment Planning chapter will lead you step-by-step through the radiotherapy planning tools and tasks in the ViewRay System™ and provide you with rationale for tasks, the background information you need to execute the tasks, the prerequisites for the tasks and examples and illustrations.

WARNING **All treatment plans must be approved by a qualified clinician before the plan can be used on a patient for radiotherapy treatment. The system operator must always be aware that the quality of the treatment plan depends on the input data. For that reason, image registration, contours, beams, dose and final plan must all be separately reviewed and each one approved by a medically qualified clinician.**

ERROR MESSAGE FOR INCORRECT DATA

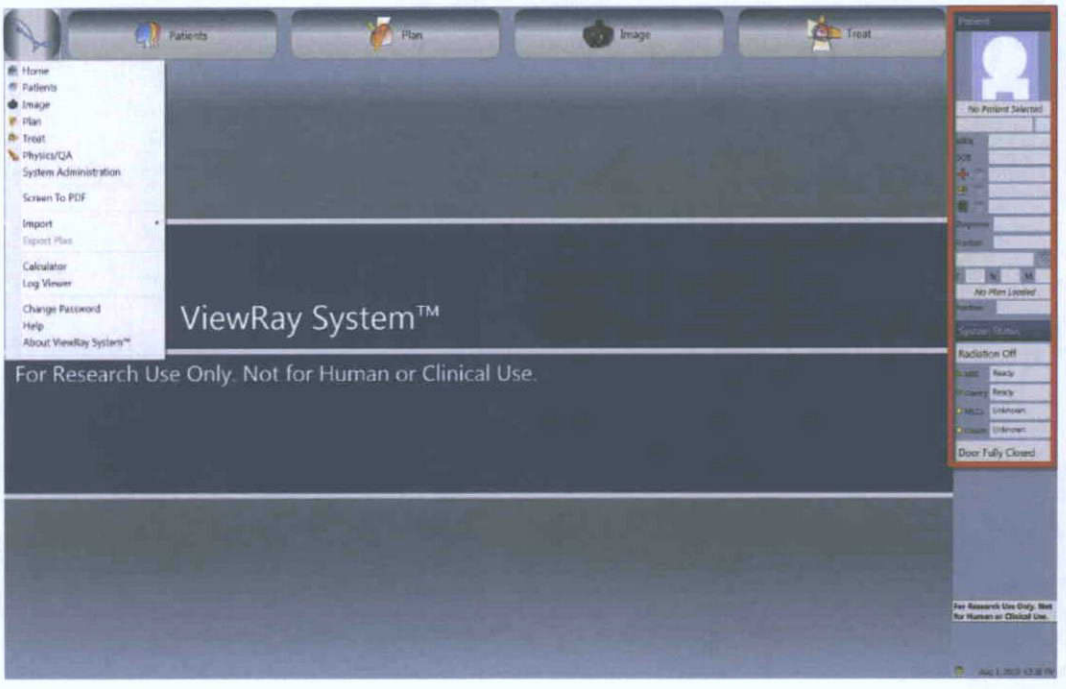


When images or patient name, ID, sex, or birthdate do not match the patient plan you have loaded, you will get this error message. The message is a warning that you are about to add information that does not match the currently loaded patient. If you want to STOP the addition of conflicting information **Click > Cancel** and the system will allow you to start another import.

Note: Clicking > OK will accept the mismatched information.

3.2 New Patient Planning

From the Treatment Planning Home Screen you can access Patients, Plan, Physics/QA, and you can export plans and import images, structures, contours, and dose.



PATIENT SIDEBAR

ALWAYS verify that the patient to be treated and the patient identified on your operator screen match.

The patient sidebar is on the right side of every screen.

The following patient information can be displayed: patient photo, patient name, gender, plan name, diagnosis, date of birth, fraction, prescription, TNM staging, and treatment position.

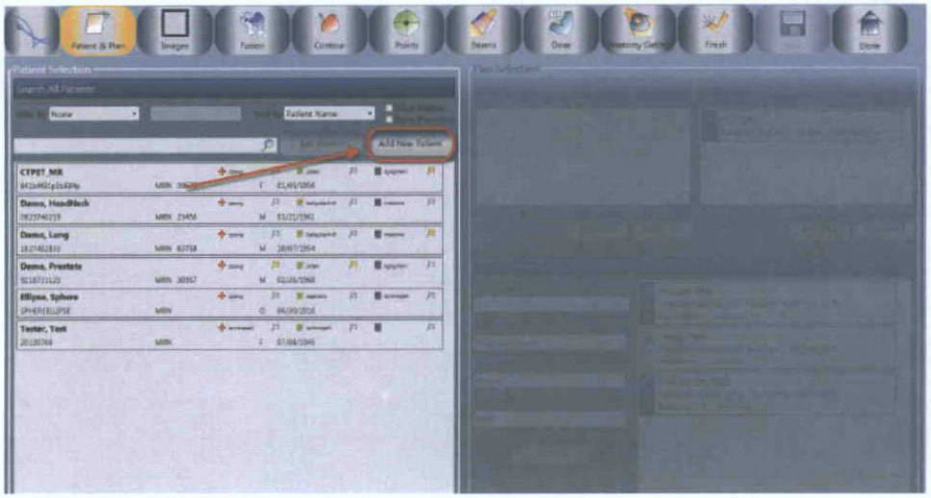
In addition, the names of the treating physician, physicist and dosimetrist are displayed on the sidebar.

Careful attention to the sidebar information can help ensure that patients are accurately identified with their plan and treatment fraction.

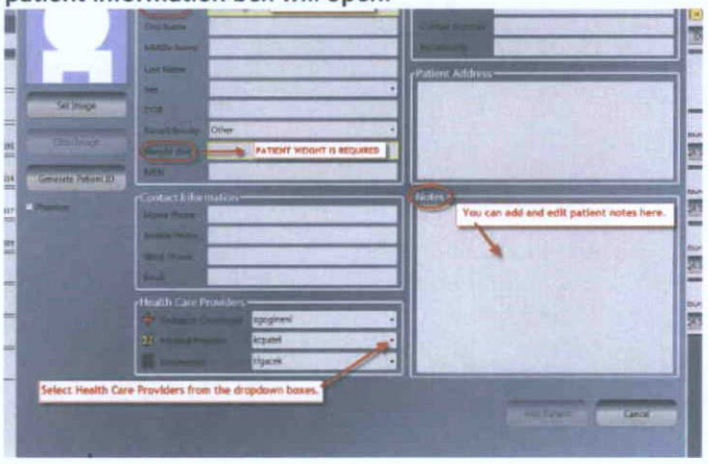
The View Ray System may be contraindicated for patients with implants due to the presence of the MRI magnetic field. All patients must be screened for MRI compatibility prior to treatment.

3.2.1 Registering a New Patient

1. From the **Home** screen **Click > Plan.**
   - ❖ The **Patient & Plan** screen will open.
   - ❖ The **Patient Selection** group box is on the left side of your screen and enabled. The right side of the screen will be unavailable (grayed out).
2. **Click > Add New Patient.**



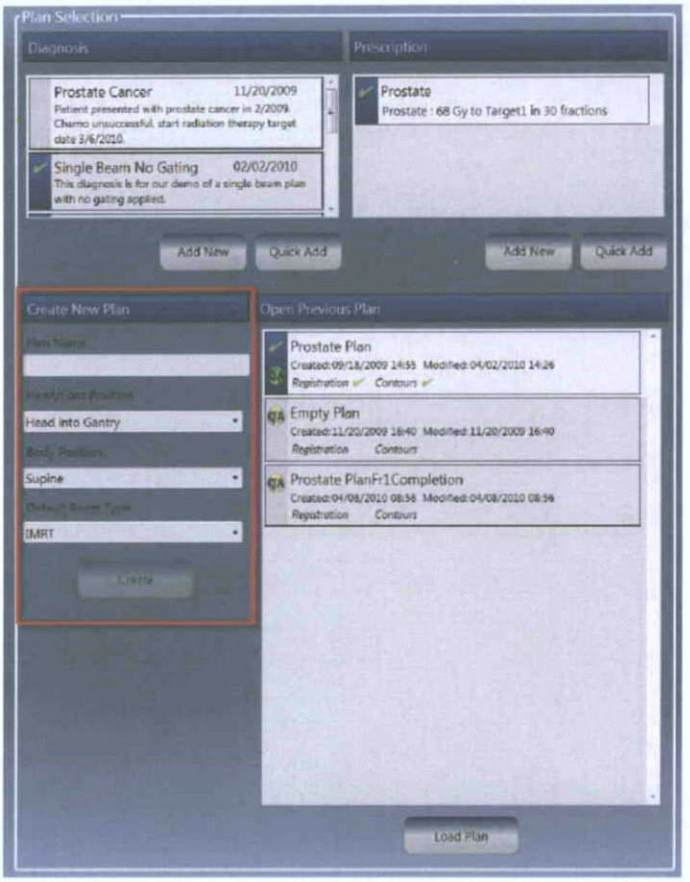
❖ A patient information box will open.



3. Enter the patient information in the boxes provided.

4. **Click > Add Patient**
   - ❖ The new patient is added to the patient list.

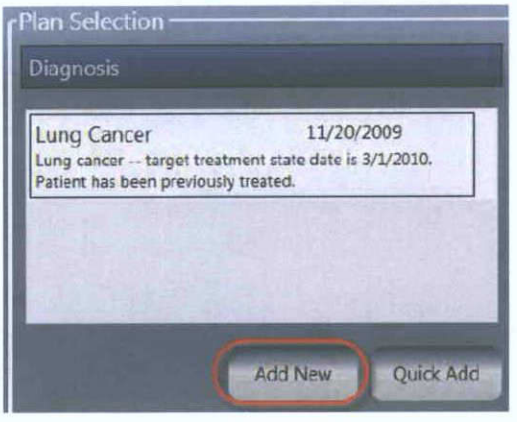
The **Plan Selection** group box is on the right side of the **Patient & Plan** screen. From this group box you can add a diagnosis, a prescription, create a new plan, open a previous plan, or load a plan.



3.2.2 Adding a Diagnosis

From the Diagnosis section of the **Plan Selection** group box

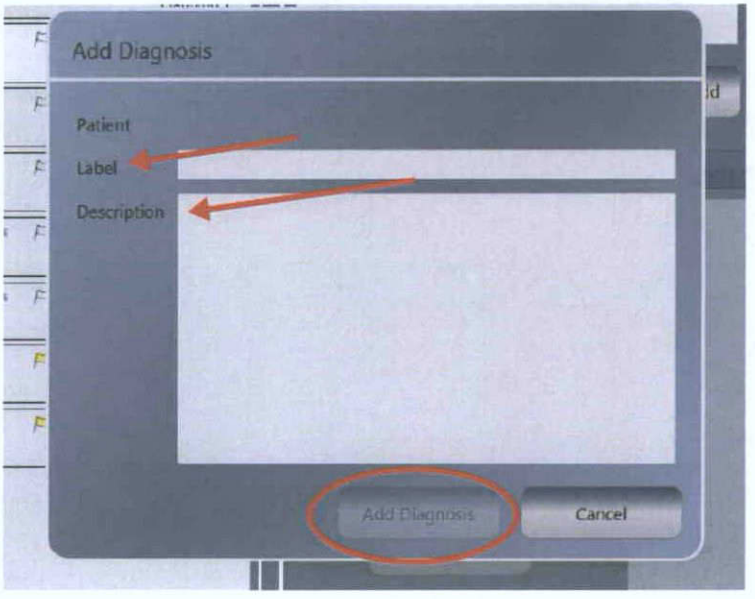
1. **Click > Add New.**



❖ An information box will open in the center of the **Patient & Plan** screen.

2. **Add the diagnosis Label and Description in the information box.**

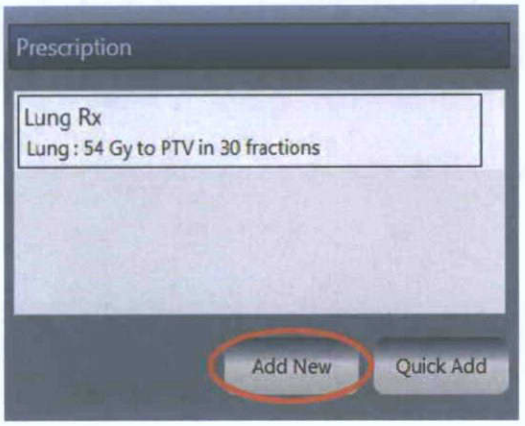
3. **Click > Add Diagnosis.**



3.2.3 Adding a Prescription

From the **Prescription** section of the **Plan Selection** group box

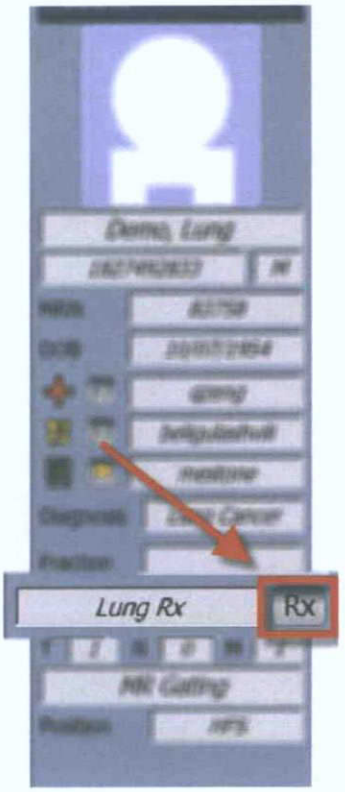
Click > Add New.



- A **Prescription** information box will open in the center of the **Patient & Plan** screen with boxes for **Prescription Label, Primary Prescription, TNM Values, Treatment Options, Fractions and Dose Constraints.**

3.2.4 Returning to Prescription

You can return to the **Prescription** information box from any screen by **Clicking the Rx button on the Patient Sidebar.**



3.2.5 Entering a Prescription Label

A new prescription must have a label. At the top left of the **Prescription information box** enter the label for the prescription.

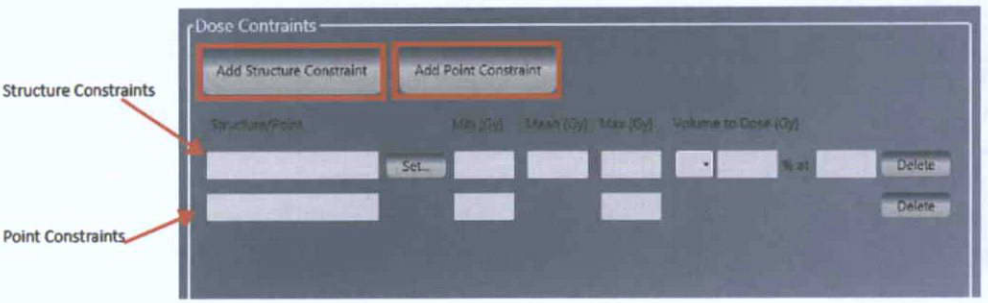
❖ The other required field on the prescription information box is the TNM site. (See Entering TNM Values on page 68).

3.2.6 Adding Structure, Dose, and Point Constraints

To add a structure constraint.

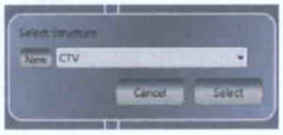
1. Click > Add Structure Constraint.

   ❖ Empty boxes will appear for the Structure name, Min, Mean and Max Grays, and Volume to Dose.



2. Click > Set

   ❖ The Select Structure dropdown box will appear in the middle of the screen.



3. Click > The dropdown arrow.

   ❖ The structure list expands.

4. Click on the structure you want from the list.

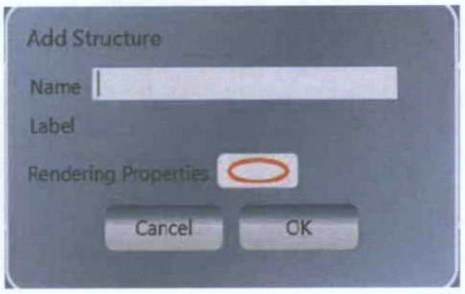
   ❖ The selected structure will appear in the information box and the list will close.

5. Click > Select.

3.2.7 Creating a New Structure

1. Click > New at left of the Select Structure information box.

❖ The Add Structure information box opens.

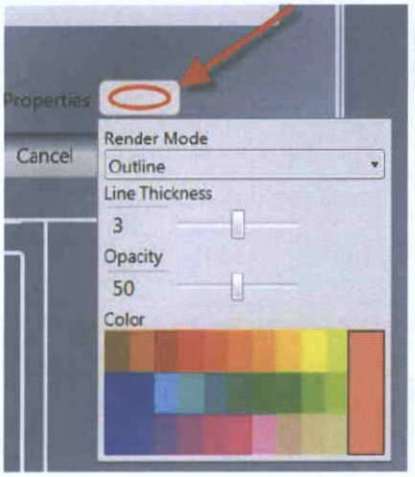


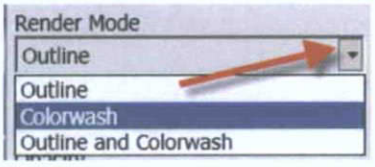
2. Add a structure name.
   ❖ The label is the same as the structure name you typed in and will be filled in automatically.

3. To set the rendering properties, Click on the red oval next to Rendering Properties.

❖ The Render Mode menu will open.

4. Click the dropdown arrow to choose from Outline, Colorwash and Outline and Colorwash.





5. You can use the sliders to select the line thickness and opacity, and the color palette to select a structure contour color.

6. When you are finished making your Render Mode choices, Click anywhere outside the menu to close the dropdown box.

3.2.8 Setting the Mean and Volume-to-Dose Constraints

1. **Click > dropdown arrow.**
   - ❖ A list of relational symbols will open.
2. **Click on the symbol you want.**

For Volume to Dose you must additional **enter the volume percent.**

4. **Enter the Grays** for your dose constraint.

3.2.9 Adding a Point Constraint

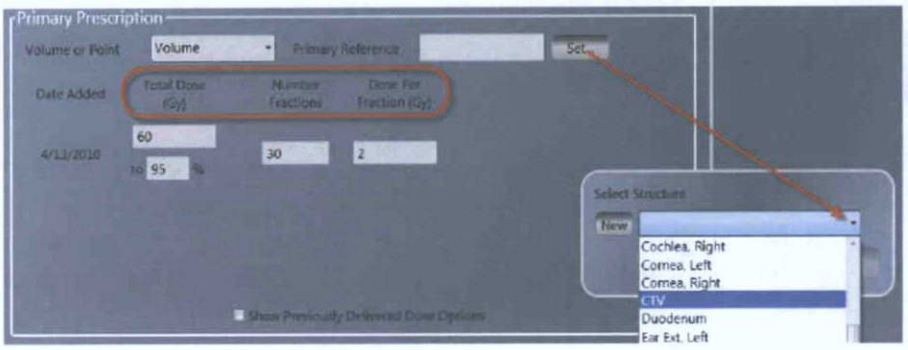
1. From the **Dose Constraints** group box **Click > Add Point Constraint.**
   - ❖ Three empty boxes will appear.

2. In the box on the left add the name of the **Point.**
3. In the middle box add the **Minimum Gray** to be delivered to the point.
4. In the last box add the **Maximum Gray** to be delivered to the point.

3.2.10 Adding a Primary Reference

You can add the primary reference to a **Volume** or a **Point.**

1. **Click > dropdown arrow** in the **Primary Prescription** information box.



2. **Select** either **Volume** or **Point** from the dropdown menu.

➢ If you select **Volume**:

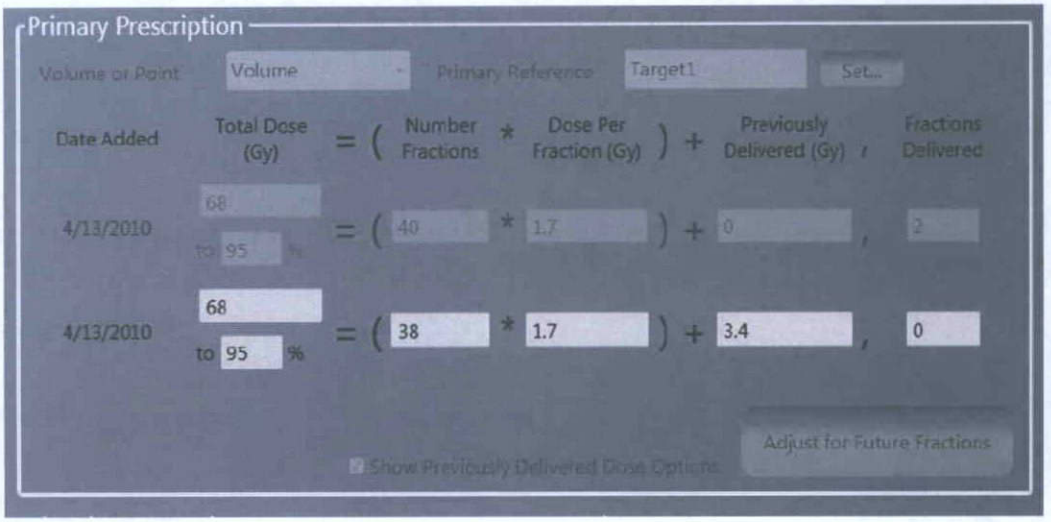
1. **Click > Set** next to the **Primary Reference** information box.
   - ❖ The **Select Structure** information box appears.
2. **Click > dropdown arrow**
   - ❖ The list of structures will expand (as shown above).
3. **Click** on the structure you want from the list.
   - ❖ The expanded list will close and the structure you selected will appear in the information box.
4. **Click > Select** (this button is on the bottom right of the **Select Structure** information box).
   - ❖ The selected structure will now appear in the **Primary Reference** information box.
5. Type in the **Dose to Volume %, Number of Fractions** and the **Dose Per Fraction is calculated.**

➢ If you select **Point:**

1. Type in the name of your **Primary Reference** (for example: Isocenter).
2. Type in the **Total Dose, Number of Fractions**, and **Dose Per Fraction.**

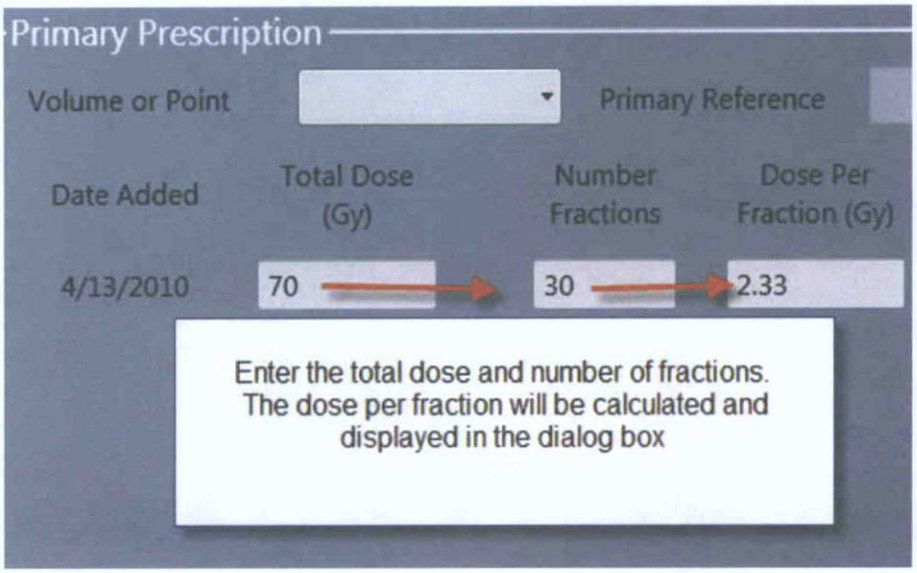
3.2.11 Adjusting for Future Fractions

This function is an adaptive treatment feature. You can be use it to calculate fractionated treatments for patients who have previously received radiation therapy or to boost remaining fractions after reviewing your treatment plan.



1. Check the **Show Previously Delivered Dose Options.**
   - ❖ Information boxes for **Total Dose, Number of Fractions, Dose Per Fraction, Previously Delivered** and **Fractions Delivered** will open.
2. Fill in the information boxes and **Click > Adjust for Future Fractions.**
   - ❖ A second set of information boxes will open with the future fractions adjusted based on the **Previously Delivered Dose.**

3.2.12 Setting Total Dose, Fractions and Dose Per Fraction



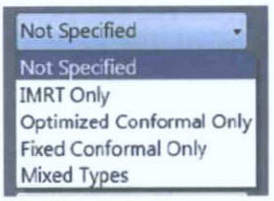
3.2.13 Setting Treatment Options

The Treatment Options you choose in the Prescription will be the *only options available* as you progress through the treatment planning and delivery screens.

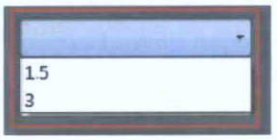
1. Click > the dropdown arrow on one of the options.

❖**The list of choices for that option will expand.**

2. **Click** on the option you want to select.



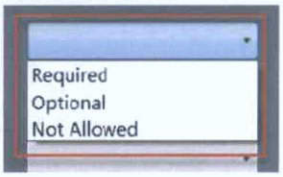
Beam Type Treatment Options



Slice Thickness Treatment Options

In-Plane Resolution Treatment Options

Setup Imaging, Imaging During Treatment, Anatomy Gating, Predict Dose, and **Reoptimize Dose** all have the same Treatment Option Choices—**Required, Optional,** or **Not Allowed.**



3.2.14 Entering TNM Values

There are dropdown boxes with a range of choices for each TNM value:

- The T *(primary **tumor**)* choices range from **T0**, (in situ) to **T4**.
- The N *(regional lymph **nodes**)* choices range from **N0** to **N3**.
- The M *(distant **metastasis**)* choices are **M0** or **M1**.
  - ➢ Grade choice is **0** to **10**.

To select your TNM values:

1. **Click** the dropdown arrow at the right of each box.
   - ❖ The value ranges will appear for each box.
2. Click the desired value.

The **Intent** dropdown box on the left has three choices: **Curative, Palliative, Other.**

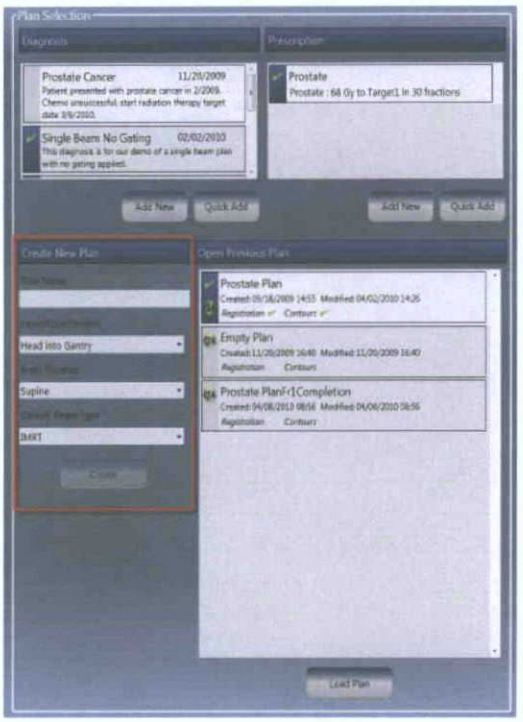
The next **Intent** dropdown box has two choices: **Malignant, Benign.**

NOTE **Site is a required field.**

3.3 Creating a New Plan

After you have added both a diagnosis and prescription for a patient you can create a plan.

1. Fill out the **Plan Selection** group box by adding a **Plan Name** and selecting **Head/ Foot Position, Body Position** and **Default Beam Type** from the dropdown boxes.



2. **Click > Create**
   - ❖ Your new plan will appear in the **Open Previous Plan** box and the **Patient ID** sidebar will be updated with the new plan name.

3.4 Selecting an Existing Patient

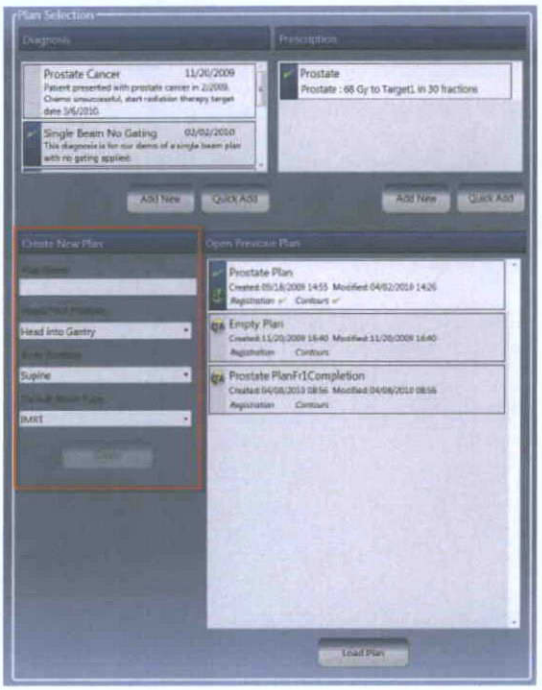
To select an existing patient go to the **Patient & Plan** screen and **Select** a patient from the **Patient Selection** group box in one of the following ways:

Click on the patient's name.

or

Use the **Search All Patients incremental search box.** Begin typing the patient's name, ID or birthdate into the incremental search box and the patient names will be continually sorted as you type.

❖ A list of the saved diagnoses, prescriptions, and plans for the patient will be displayed.

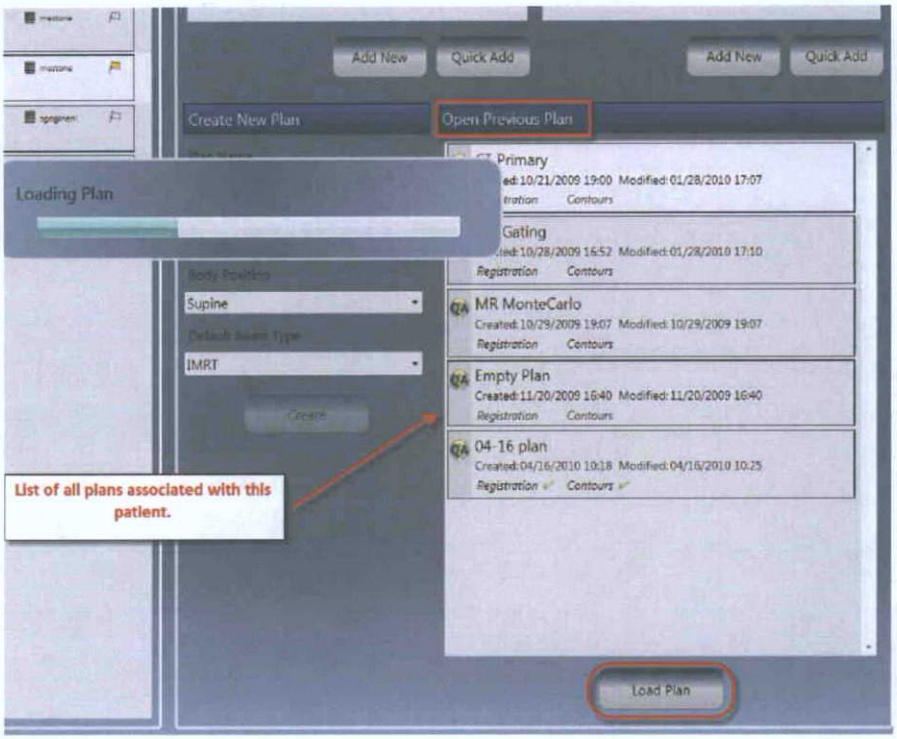


You can load a previous treatment plan, select or add a diagnosis, select or add a prescription, and create a new plan from the **Plan Selection** group box on the **Patient & Plan** screen.

3.5 Loading a Plan

To load a plan **Select** it from the **Open Previous Plan** list by **Double-clicking** on the plan name or highlight the plan name by **Clicking** on it once and then **Click > Load Plan.**

- A progress bar will appear as the plan loads. When the progress bar disappears your plan has loaded and you can continue in the treatment planning workflow.



3.6 Images

3.6.1 Importing Images to a Plan

A plan must be loaded before you can import images. After loading a plan go to the **Images** screen.

1. Click > **Images** from the workflow bar.

❖ The **Images** screen will open.

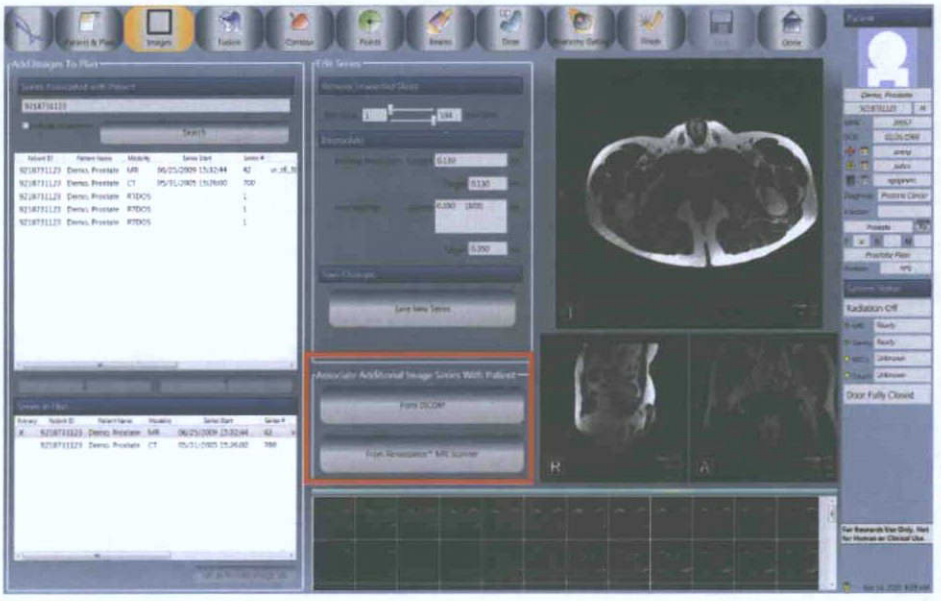
If images are already associated with the patient they will be listed in **Series Associated with Patient.**

To import a series from this list:

2. **Highlight** the series and **Click** > the **Down Arrow** below the series list.
   ❖ The series you choose first becomes the **primary series** and will have an X before the series name.

To change the **primary series**:

1. **Highlight** the series you want to set as primary.
2. **Click** > **Set as Primary Image Set.**



At the bottom middle of the **Images** screen is the **Associate Additional Image Series With Patient** group box.

Start an image import from the Associate Addition Image Series with Patient group box in one of two ways:

- Click > From DICOM
  - ❖ An information box will appear and you will be prompted to select the import Provider.

3.6.2 Removing Slices

You can remove slices from the beginning or end of a series by using the sliders at the top right of the **Images** screen. When you have finished selecting the slices you want removed:

Click > Save New Series.

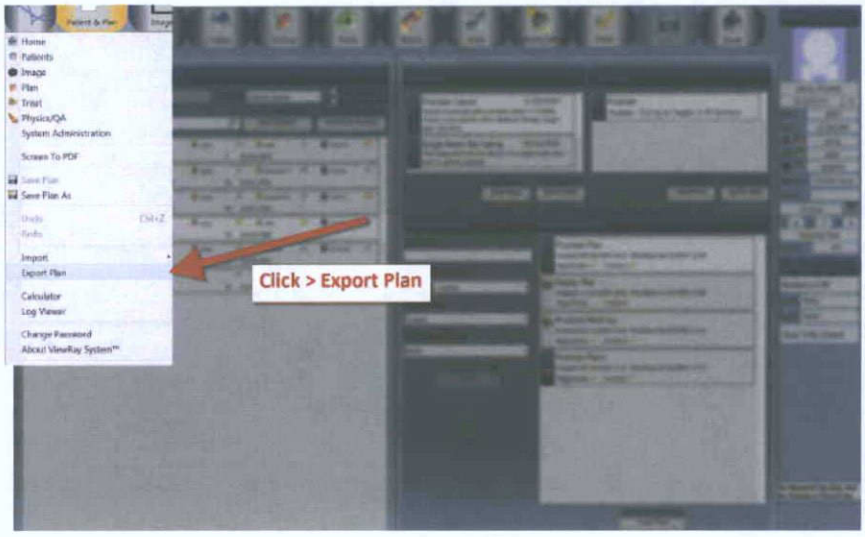
- ❖ The edited series will be listed in the **Series Associated with Patient** box and identified as "Edited".

3.6.3 Exporting ViewRay System Plan Data via DICOM

To export a plan or any plan data you must first load the plan. See Loading a Plan on page 71.

After loading a plan:

1. **Click > ViewRay Icon.**
   - ❖ Dropdown menu appears.
2. **Click > Export Plan.**

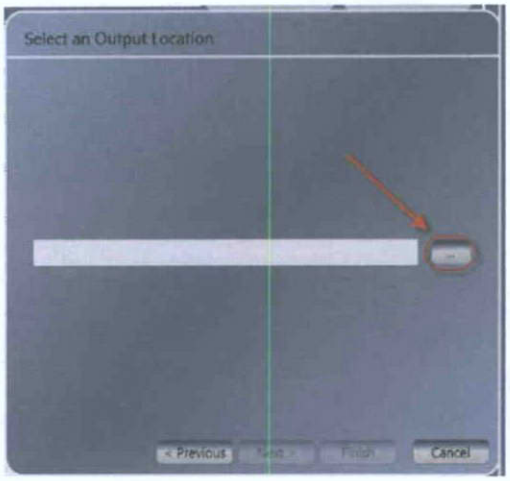


- ❖ The **Export Plan** information box will open.

3. Check the box next to the parts of the plan you want to export.
4. **Click > Next.**

❖ The **Select an Output Location** information box will open.

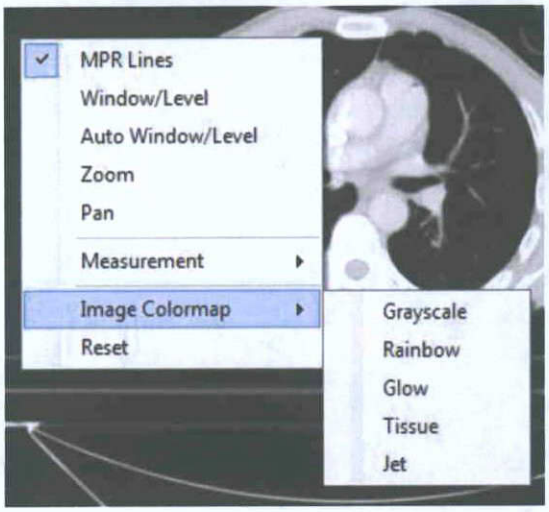
1. Click > the **command button** to open a menu list of output locations.
2. Choose a location from the list.
3. Click > **Next.**



❖ The plan will begin exporting. When the export is finished, the **Export Complete** information box will open.

4. **Click > Finish.**

3.6.4 Viewing Images



By **right clicking** the mouse anywhere on an image you access the image options for each screen.

The available selections will change depending on the screen.

When viewing registered volumes you can adjust the method of fused viewing, such as red intensities for PET data and standard grayscale for MRI data.

Clicking the mouse wheel will toggle between a single image view and "four up".

You can display any slice in axial, lateral or vertical orientations. The ViewRay System™ accounts for patient orientation when displaying anatomical markers for left/right.

To zoom in on the 3D volume, **Click the mouse wheel on the volume image and move the wheel forward (away from you).** To zoom out, **Click the mouse wheel on the volume image and move the wheel backward (toward you).**

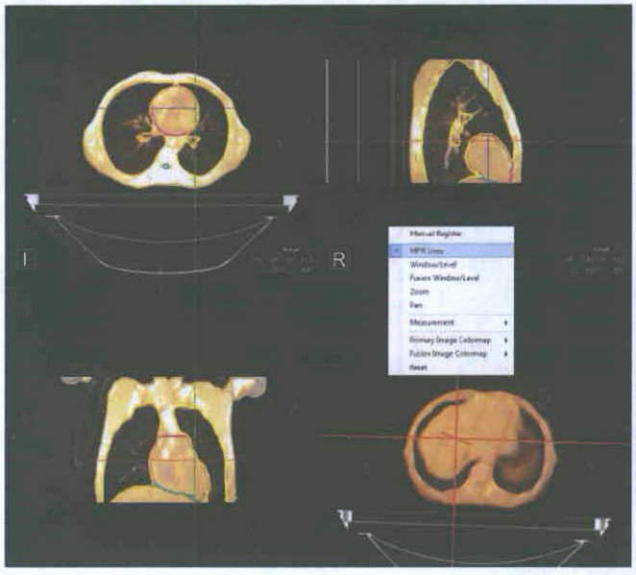
3.6.5 Using MPR Lines

Multiplanar lines allow you to view slices in vertical, axial or lateral planes. The coordinate planes are based on the ViewRay System™ machine coordinates. See ViewRay System™ Coordinates on page 7.

Moving the green line with your mouse moves through the slices in a lateral plane.

Moving the blue line with your mouse moves through the slices in a vertical plane.

Moving the red line with your mouse moves through the slices in an axial plane.



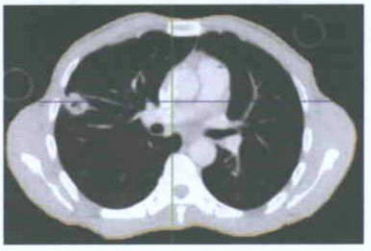
3.6.6 Using Window/Level Settings

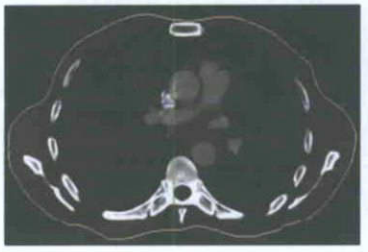
Window and **Level** settings adjust the contrast and brightness of your images.

Window defines the upper and lower limits of the range of pixel values. Setting a window level determines which values will be seen as white and which as black.

A large dynamic range of window settings has a broader range of values between the pixel value defined as black and the pixel value defined as white and therefore has greater contrast—a whiter white area and blacker black areas. This high contrast helps define borders but can mask fine details.

Level provides contrast enhancement of a specific part of the signal range by isolating the values of interest and creating a contrasted area between those values. **Level settings distinguish and define regions of close radiographic densities.**





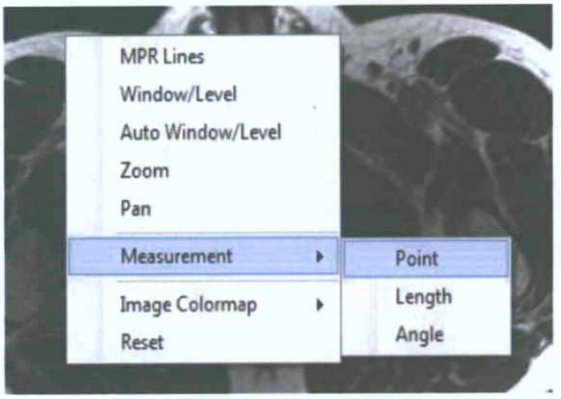
3.6.7 Using the Zoom tool

1. **Right Click** anywhere on the image.
   - ❖ The viewing tools will appear.
2. **Click > Zoom.**
3. **Click and drag your mouse forward** (away from you vertically) to Zoom in.
4. **Click and drag your mouse back** (toward you) to Zoom out.

3.6.8 Using the Pan tool

1. **Right Click** anywhere on the image.
   - ❖ The viewing tools will appear.
2. **Click > Pan.**
3. **Click and drag your mouse in any direction.**

3.6.9 Using the Measurement Tools



You can use the measurement tools to calculate the density and coordinates of any point, to measure the length of any area or structure, and the angle of any two lines.

3.6.10 Using the Point tool

1. **Right Click** anywhere on the image.
   - ❖The viewing tools will appear.
2. **Click > Measurement.**
   - ❖ The sub-menu will appear with three selections—Point, Length, Angle.
3. **Click > Point.**
4. **Click** anywhere on the image to designate a Point.
   - ❖The density and coordinates of the point you selected will be displayed at the top left corner of the image.

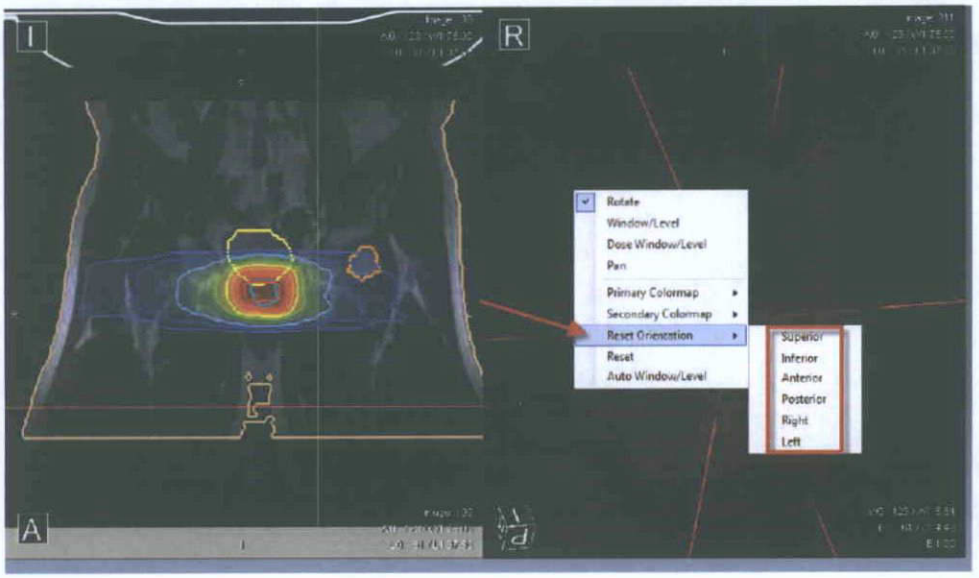
3.6.11 Using the Length tool

1. **Right Click** anywhere on the image.
   - ❖The viewing tools will appear.
2. **Click > Measurement.**
   - ❖The sub-menu will appear with three selections—Point, Length, Angle.
2. **Click > Length.**
3. **Click** a beginning point on the image and **drag** to the end of the area you want to measure.
   - ❖The length of the line you drew will appear at the top left corner of the image.

3.6.12 Image Orientation Reset

To reset the image orientation:

1. Right Click on the image.
   - ❖ The image setting box will appear.
2. **Click > dropdown arrow** next to **Reset Orientation.**
   - ❖ The orientation selections will expand.
3. Click the orientation you want to select.



3.7 Using the Display Options

The **Display** options are available on the **Fusion, Contour, Points,** and **Beams** screens.

After a plan and images have been loaded you can use the **Display** tools to change the viewing options for **Image View, Structures, Points of Interest, Beams,** and **Dose View.**

3.8 Removing Overlays

Structures, Points of Interest, Beams, and Dose View are all image overlays and can be turned off.

*Use the **View None** button to turn one or all of the overlays off.*

3.9 Editing Image Sets

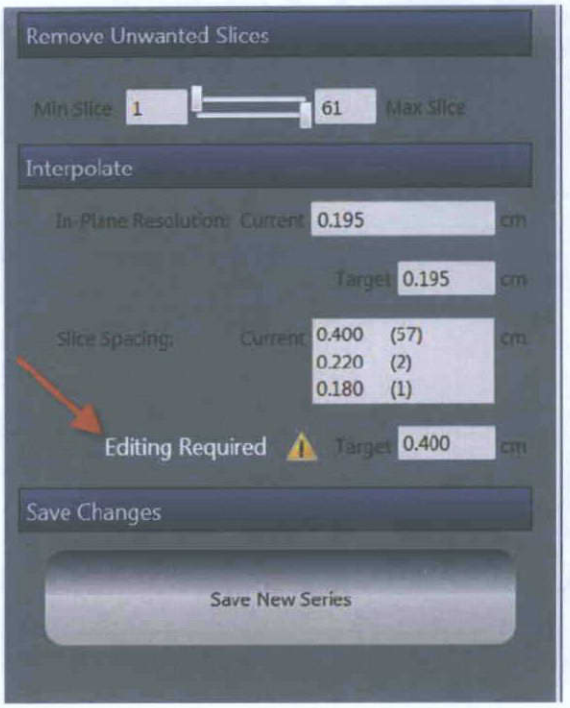
You can **Remove Unwanted Slices** from the beginning and end of an image series by using the sliders to indicate the desired number of removed slices, or you can type in the numbers.

Interpolating images by modifying the voxels (volume elements) across-slice resolution to match the within-slice resolution clarifies details and 3D volume visualization.

You must interpolate images when the **Editing Required** warning is displayed.

To interpolate images:

1. Choose target sizes for both the **In-Plane Resolution** and **Slice Spacing**.
2. Click > **Save New Series**.



3.10 Image Registration

You can fuse MRI, CT and PET volumes.

3.10.1 Registration Modes

There are three modes of registration:

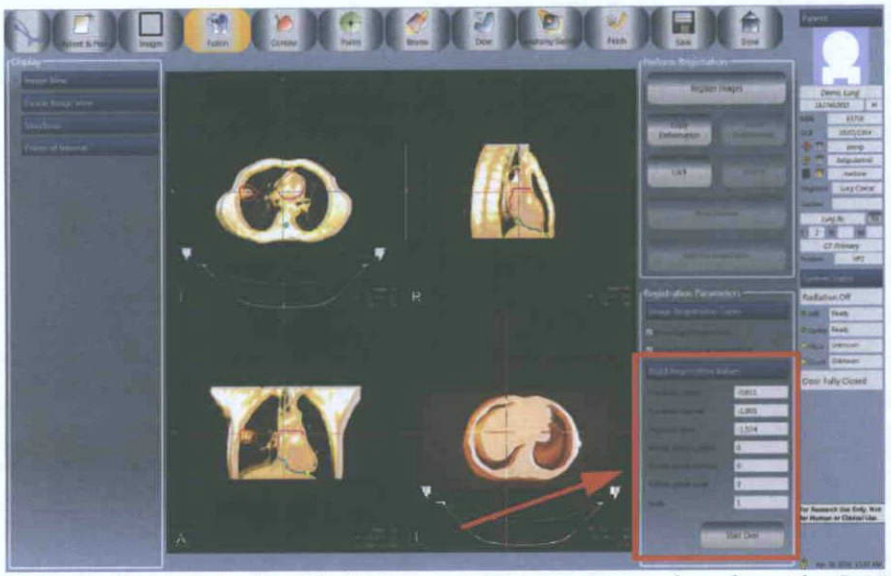
Manual Rigid Registration allows you to shift the secondary volume until it aligns with the primary volume.

Automatic Rigid Registration uses a software algorithm to align the volumes.

Automatic Deformable Registration uses an intensity-based software algorithm and compares the intensity patterns in images using correlation metrics to deform the secondary volume to match the anatomy in the primary image set.

❖ **After choosing your primary series and at least one other series from the Image screen, you can register images.**

Allow Rigid Registration and **Allow Deformable Registration** are checked by default. Rigid Registration runs first followed by Deformable Registration.



You can also manipulate the images using registration values from the **Registration Parameters** box located at the lower left of the Fusion screen.

There are three methods to enter the values:

You can **key in the values** you want or **use the mouse wheel** over the registration value boxes to move values up or down, or **drag the secondary image set to align with the primary image set** and the registration values will change to correspond with the image registration.

After using one of those methods to align the two image sets,

Click > Register Images.

- A progress bar will appear at the bottom of the patient sidebar. When the progress bar disappears, the image has been fused.

You can continue to register images.
or
If you are satisfied with the image registration:

Click > Lock.

If you decide you want to change the registration after you have locked it:

Click > Unlock.

To move to the next unlocked volume **Click > Next Volume.**

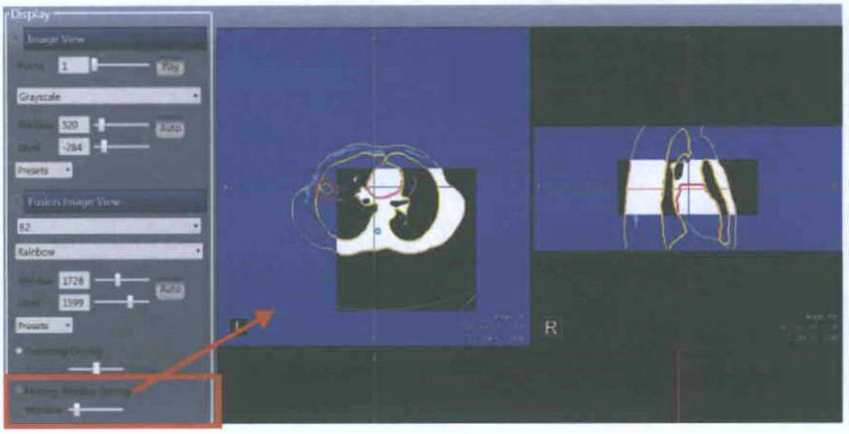
NOTE **Image registration must be approved by an authorized user.** Each image approval or unapproval requires authentication. All authentications are recorded by the system.

3.10.2 Displaying Volumes

During fusion there are two modes for displaying the primary and secondary volumes:

Colormap Overlay (default mode) displays the secondary image set fused with the primary image set.

Moving Window Overlay (shown below) displays the primary image set in a rectangle that can be dragged and moved overtop the secondary image set.



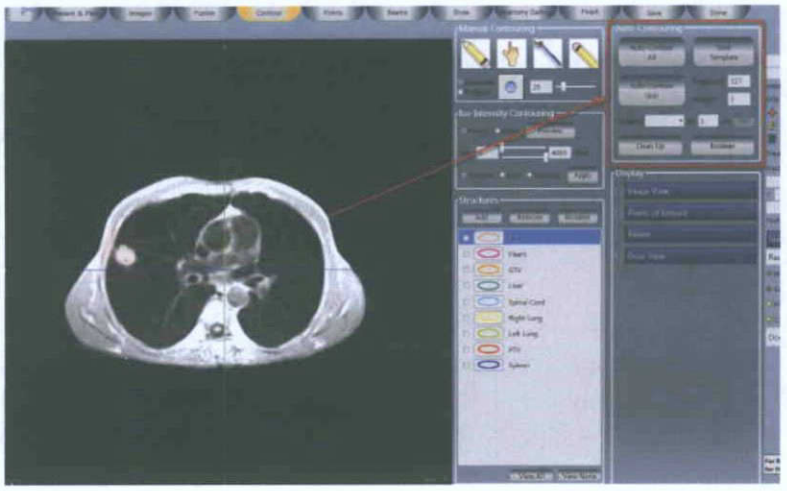
3.11 Contouring

The accuracy of your patients' anatomy contours affects both the dose and the placement of the treatment fields. There are several tools in the ViewRay Treatment Planning software to assist you in the process of defining and refining the anatomy contours of your patients.

3.11.1 Auto-Contouring Using an Existing Template

Auto-contouring automatically defines the boundaries of selected anatomical structures.

1. From the **Contour** screen **Select > Auto-Contour All**



❖This will retrieve the list of stored auto-contour templates.

➢ The list includes the names of the users who created templates, template names, region of the body for each template, patient gender, and list of structures.

2. **Select** the template you want by **Clicking** on it from the list.

3. **Select > Auto-Contour**

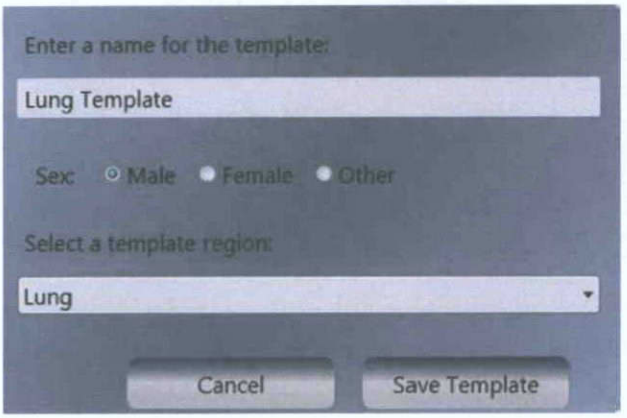
❖ A progress bar will appear in the lower right corner of the screen. Once the operation completes, the progress bar will close and all the contours that were selected in the template will be contoured on the image set.

3.11.2 Saving Auto-Contouring Templates

Once you have contoured a structure you can create your own contouring template and save it by using the **Save Template** option.

To save a newly contoured structure as a template:

1. **Click > Save Template.**
   - ❖ An information box will appear. Fill in the template name, patient gender and anatomical region. When you have finished adding the information,
2. **Click > Save Template.**



- ❖ The saved template will now be part of your template library.

CAUTION: Mislabeling a structure can lead to mistakes in setting electron density and errors in dose optimization.

3.11.3 Adding Structures

You can define and add structures using the **Structures** tool.

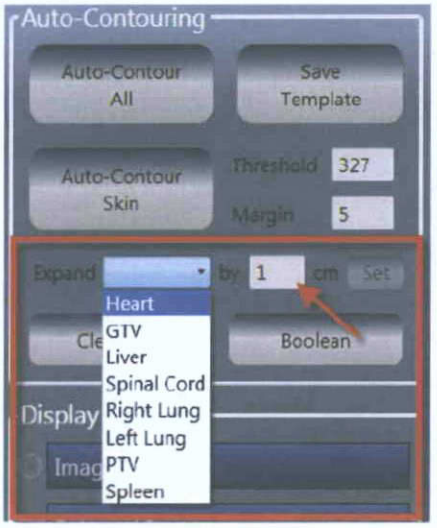
The **Structures** tool is at the center left of the image on the **Contour** screen.

1. **Click > Add.**
   - ❖ An information box will appear.
2. **Click > Select Structure.**
   - ❖ A dropdown list of structures will appear.
3. **Select a structure from the list.**
   - ❖ To customize your contour, you can choose to rename a structure after you have selected it from the list.

3.11.4 Expanding Structures

To expand a structure.

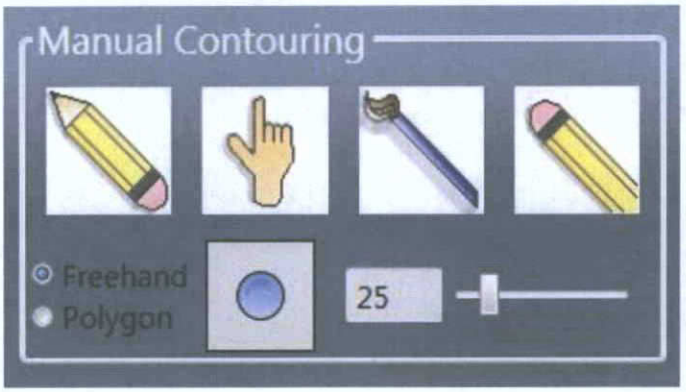
1. Click the dropdown arrow next to Expand.
   - ❖ The list of structures expands.
2. **Click the structure from the list and choose the centimeter expansion you want.**
3. **Click > Set.**



3.11.5 Using the Manual Contouring Tools

Manual Contouring has the following 4 tools:

- **Pencil** may be selected as a freehand (click and drag with autocomplete) or polygon (click on vertices and click first vertice to close).
- **Nudge** adds to the contour when starting from the inside and erases from the contour when you start from the outside.
- **Paintbrush** draws based on the size of the slider.
- **Eraser** removes contours based on the size of the slider.



NOTE: The manual contouring tools operate only on the structure you have selected from the **Structures** list. *If you are trying to use one of the tools and nothing is happening, check to make sure that the structure you are trying to contour has been **selected** from the **Structures** list.*

3.11.6 Using the Clean Up Tools

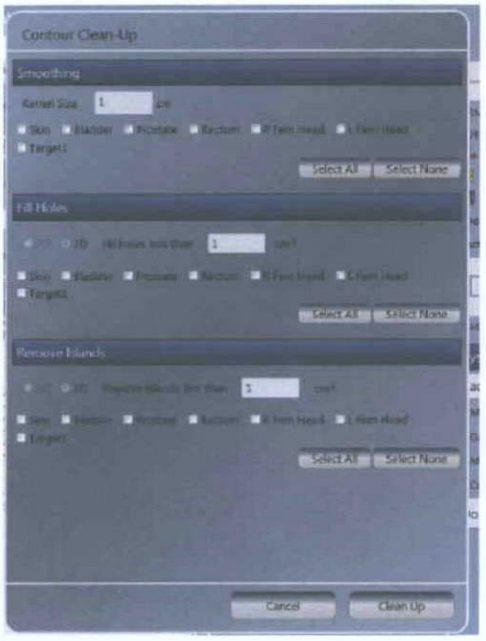
The Clean Up tools fill holes, remove islands and smooth contour lines.

To use the Clean Up Tools:

1. **Click > Clean Up** located on the right of the **Image** screen in the **Auto-Contouring box.**

❖ **The Contour Clean-up** information box opens.

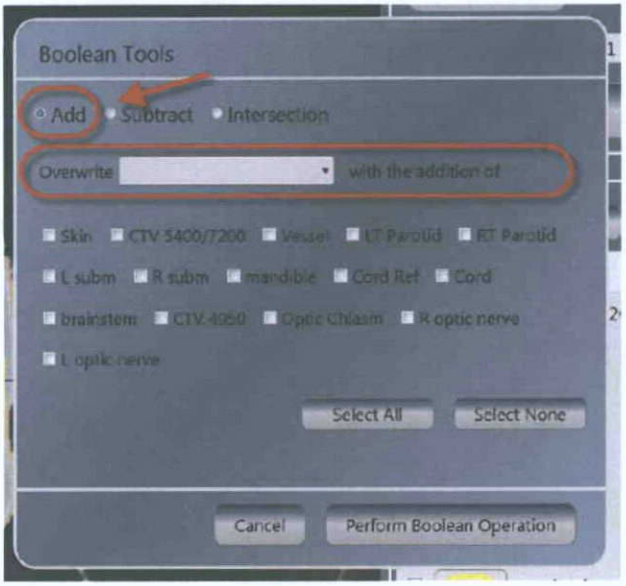
2. **Check the boxes next to the structures you want to clean up** or **Click> Select All** and specify the volume threshold in centimeters cubed for the size of the holes you want filled and islands you want removed. Set the kernel size for contour smoothing.



3.11.7 Using the Boolean Tool

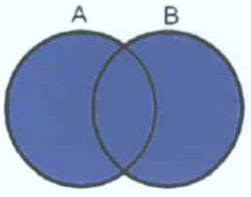
1. Click > **Boolean** from the Auto-Contouring group box.
   - ❖ The **Boolean Tools** information box will open in the center of the Contour screen.

You can use the **Boolean Tools** to overwrite (change) a structure by **adding** or **subtracting** structures to it, or by using only the **intersecting** areas of one or more adjacent structure(s).



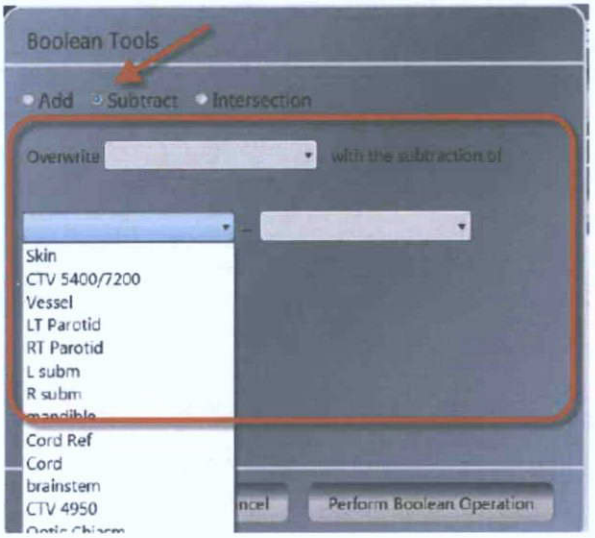
For example:

ADD—If you select the **Add** button and select one structure to overwrite from the **Overwrite** dropdown list and one or more structures from the structure list on the information box you will overwrite the structure you selected by creating a structure that is the union of the previously separate structures.

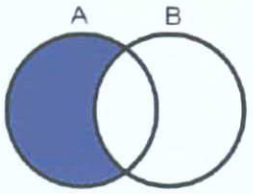


SUBTRACT—If you select the **Subtract** button, the information box will have three dropdown lists.

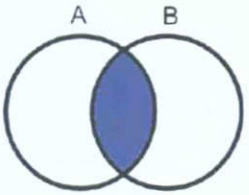
1. Select a structure you want to overwrite.
2. Select the structure you want subtracted from another.



You will create a structure that excludes the portion of the structure you want subtracted.



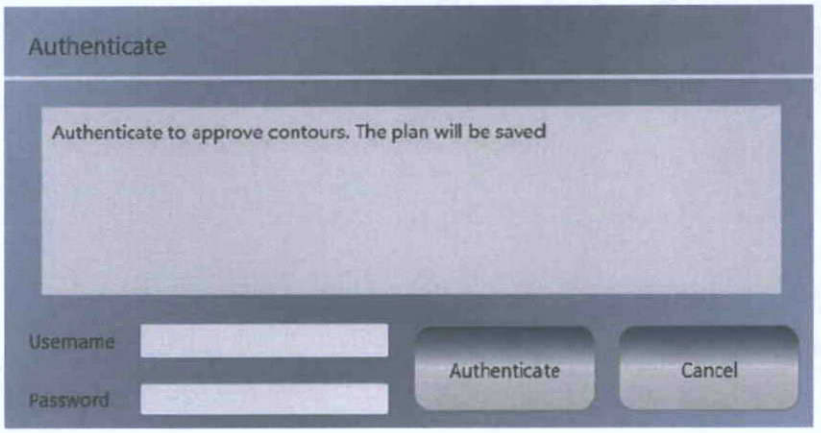
INTERSECTION—If you select the **Intersection** button and select another structure, you will create a structure that consists of ONLY the area in which the two (or more) selected structures intersect.



3.11.8 Approving Contours

When you are satisfied with the contours, go to the bottom center of the screen and **Click > Approve Contour.**

❖ The **Authenticate** information box will appear in the center of the screen.

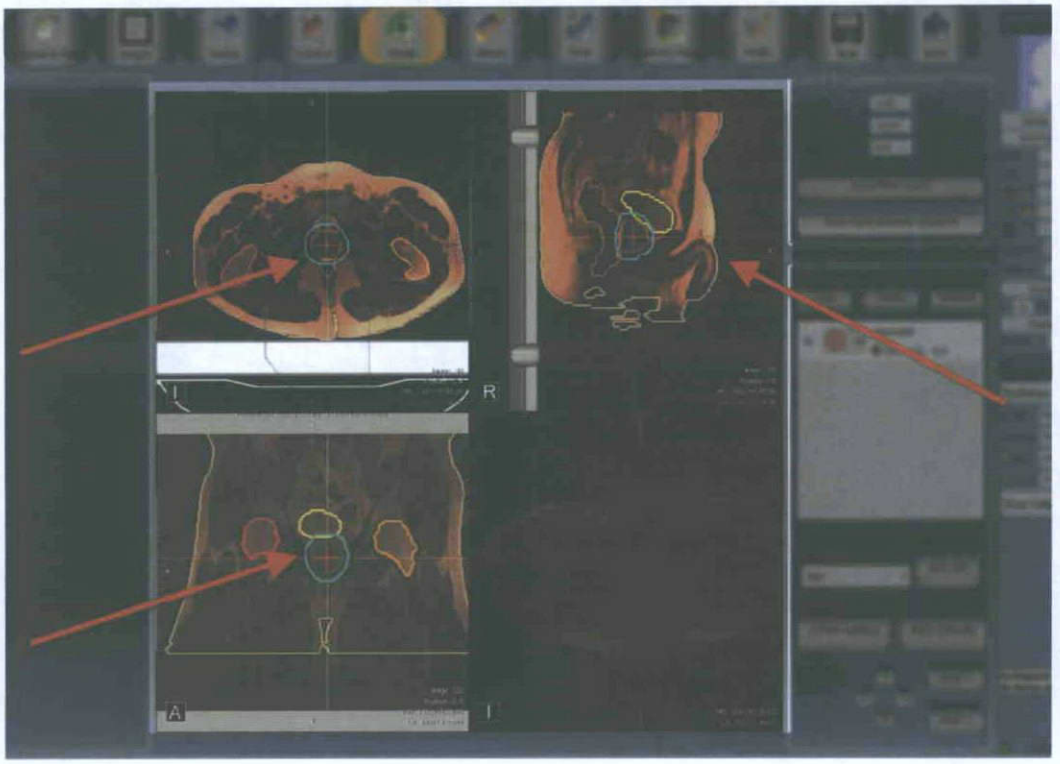


Both a username and a password are required to approve contours.

NOTE **Plans cannot be approved or delivered without approved contours.**

3.12 Points

When you open the **Points** screen, the system automatically has an **Isocenter** point placed at the center of your image volume.



3.12.1 Adding Points

To add new points:

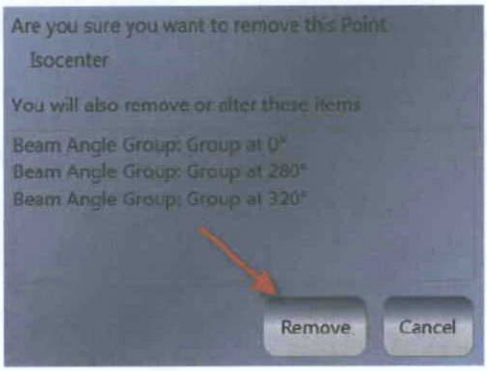
1. **Click > Add** from the **Point Placement** group box.
   - ❖ An information box will appear in the center of the screen.
2. Type in a name for the added point
3. **Click > Set.**
   - ❖ The point will be placed in the center of the image volume.

You can adjust the point's location by clicking on it and dragging it to where you want it.

3.12.2 Deleting Points

To delete points:

1. **Click > Delete** from the **Point Placement** group box.
   - ❖ A confirmation box will appear prompting you to verify removal of the point.

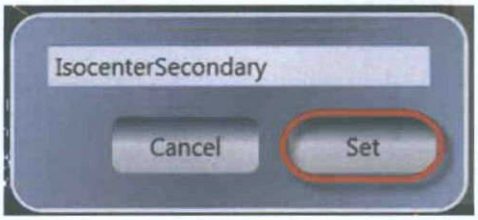


2. **Click > Remove**

3.12.3 Renaming a Point

To rename a point:

1. **Highlight** the point.
2. **Click > Rename** from the **Point Placement** group box.
   - ❖ An information box will appear in the center of the screen.
3. Key in the new name.
4. **Click > Set.**



3.12.4 Setting Points at the Center of a Structure

You can set points automatically at the center of any contoured structure by selecting the structure from the **Set at Center of** dropdown box and then clicking the **Auto-Set** button.

To manually place points:

1. **Click > Place Manually** and then click on the point and drag the point where you want it. Lock it in place by releasing the mouse.

3.12.5 Using 3-Point Landing

The **3-Point Landing** tool places a point on the patient image based on the fiducials that are visible in the image.

1. Make sure you are in the Axial view.
2. **Click > 3-Point Landing.**
3. **Click on a point on the image** based on fiducial placement and hold down the mouse and release at the point you want to end the first line.
   - ❖ A dashed red line will be drawn
4. **Click on a point arising from the first line .**
   - ❖ The first line you drew and the vertical line coming from that line will be triangulated.

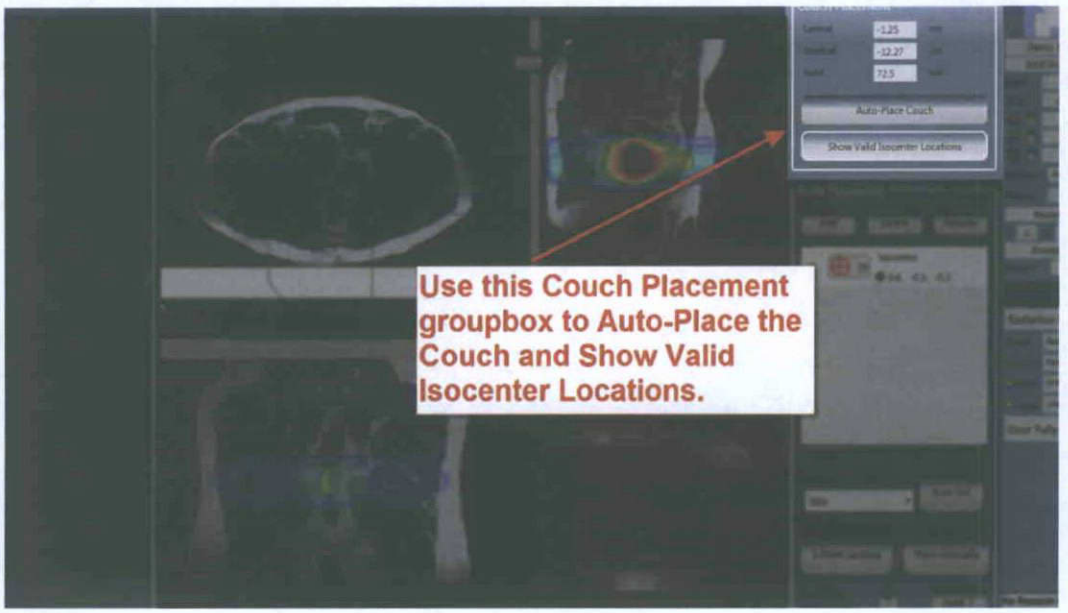
3.12.6 Fine tuning point placement

- ❑ To move one pixel at a time, **use the directional arrows** on the point placement box.
- ❑ To move 10 pixels at a time, **hold down the Ctrl key when using the arrow keys.**

3.13 Couch and Isocenter Placement

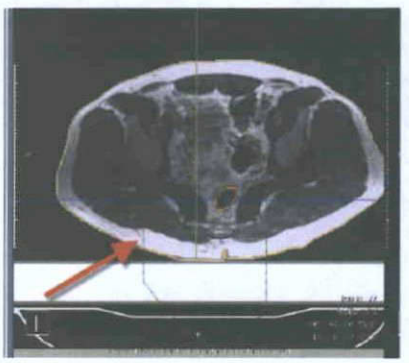
The patient's treatment planning position defines the relationship between the patient and the radiation beams. The ability to accurately reproduce the patient's planning position on a daily basis is crucial for the precise delivery of each treatment fraction.

Couch placement is an important aspect of this patient-beam relationship. Use the **Couch Placement** feature to automatically find a couch position in which the center of the image volume is placed between two table notches and where the skin contour's lowest point does not intersect with the couch.



Use the **Couch Placement** feature to show the vaild isocenter region. A valid isocenter region is an area of safe point placement that will not cause a collision of the patient or couch with the gantry.

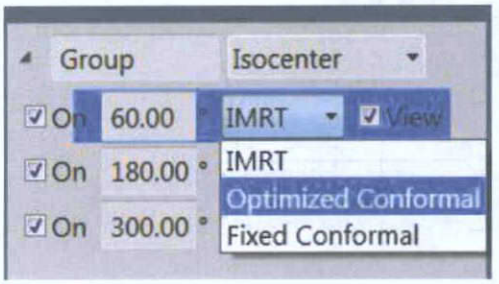
When you choose **Show Valid Isocenters** a dashed green polygon is placed around the area in which a valid isocenter can be placed.



3.14 Beams

The ViewRay System™ has three radiation beam types:

- **IMRT**
- **Optimized Conformal**
- **Fixed Conformal**



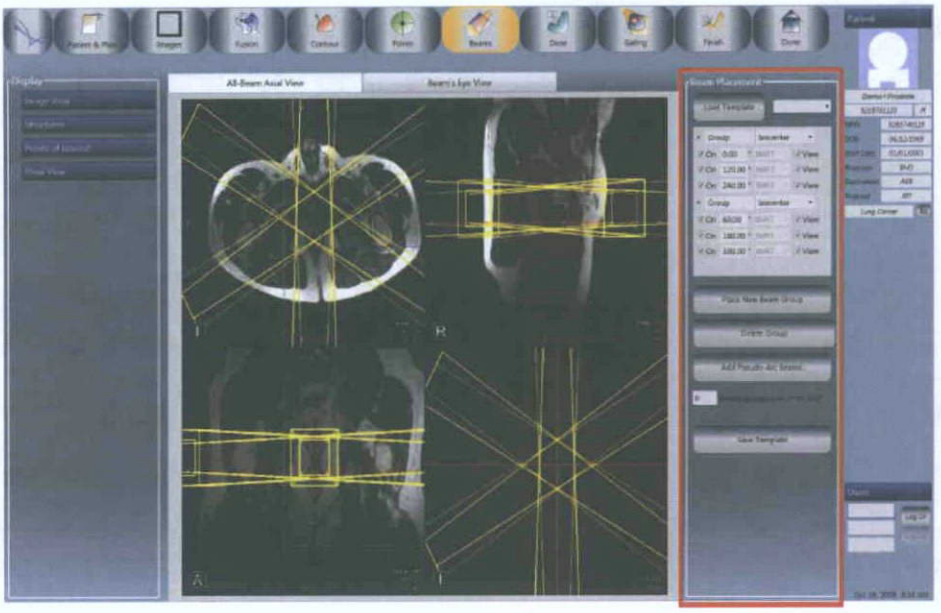
3.14.1 Beam Types Overview

IMRT: Intensity Modulated Radiation Therapy separates the radiation into segments of intensity (fluence) modulated beamlets. Each segment consists of beamlets that are shaped and their fluence modified by passing through a multileaf collimator. To use IMRT beams, you start with the desired outcome parameters (inversely plan) by entering the patient's prescription and setting the beam angles and number of beam-on segments—then choose Optimize Dose and the treatment planning system calculates the MLC shapes and fluences for the beam-on segments.

Optimized Conformal: Optimized Conformal beams consist of a single segment shaped to fit the target. The beams maintain their one shape during the entire fractionated treatment. The treatment planning system uses operator-chosen constraints to calculate and set the fluence of optimized conformal beams. Because these beams are shaped to conform to the tumor volume, the radiation to the surrounding normal tissues is restricted.

Fixed Conformal: Both the shape and angle of fixed conformal beams are operator-selected. The treatment planning system calculates and sets the beam fluence based on the prescription weighting points.

The **Beam Placement** group box is on the right of the **Beam** screen. From this group box you can load beam templates, place new beam groups, delete beam groups, add pseudo-arc beams and save beam templates.



The beams you add will default to the beam type selected when the plan was created. You can change the beam type by clicking on the arrow next to a beam to open the dropdown menu and then selecting a different beam type.

3.14.2 Placing a Beam Group

1. Click > Place New Beam Group
   - ❖ A group of three beams will be placed at the isocenter.

You can change both the isocenter and the beam type for each beam in a group.

To change the beam type:

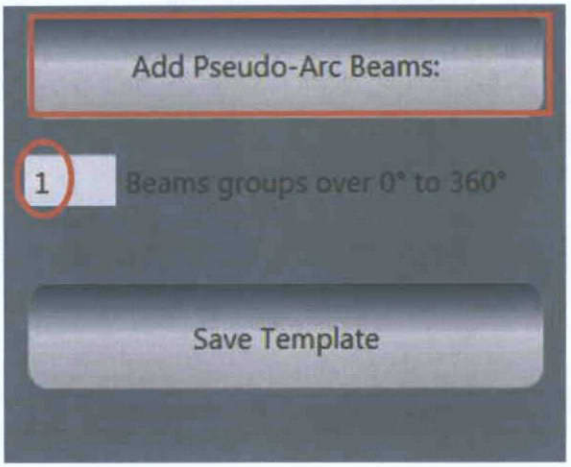
1. Click > on the Group you want.
2. Double-click on a beam to Select it and open the dropdown box of beam types.

To change the isocenter:

1. Click > on the dropdown arrow next to Isocenter.

3.14.3 Adding Pseudo-Arc Beams

1. Enter in the text box the number of beam groups you want to add.
2. Click > Add Pseudo-Arc Beams.
   - ❖ The beam groups will be added and spaced evenly over a 0° to 360°field.



3.14.4 Manually Adjusting Beam Angles

1. Select a beam group and Click and hold down on one of the beams to drag and rotate the beam group.
2. Release the Click to lock the new beam group placement.

3.14.5 Adjusting Beam Angles Using Beam Values

1. In one of the individual beam boxes of a group of beams type in a beam angle.

❖The other beam angles in the group will also change so that all beams are spaced evenly over a 0° to 360°field.

3.14.6 Loading Beams from Templates

Use the dropdown box to view and select a beams template.

3.14.7 Deleting Beam Groups

Beam groups can be deleted by clicking on the Delete Group button. Each click deletes the first beam group in the list.

If you want to delete a different beam group:

1. Highlight the beam group by clicking on a beam from the group.
2. Click > Delete Group.

3.14.8 Saving Beam Templates

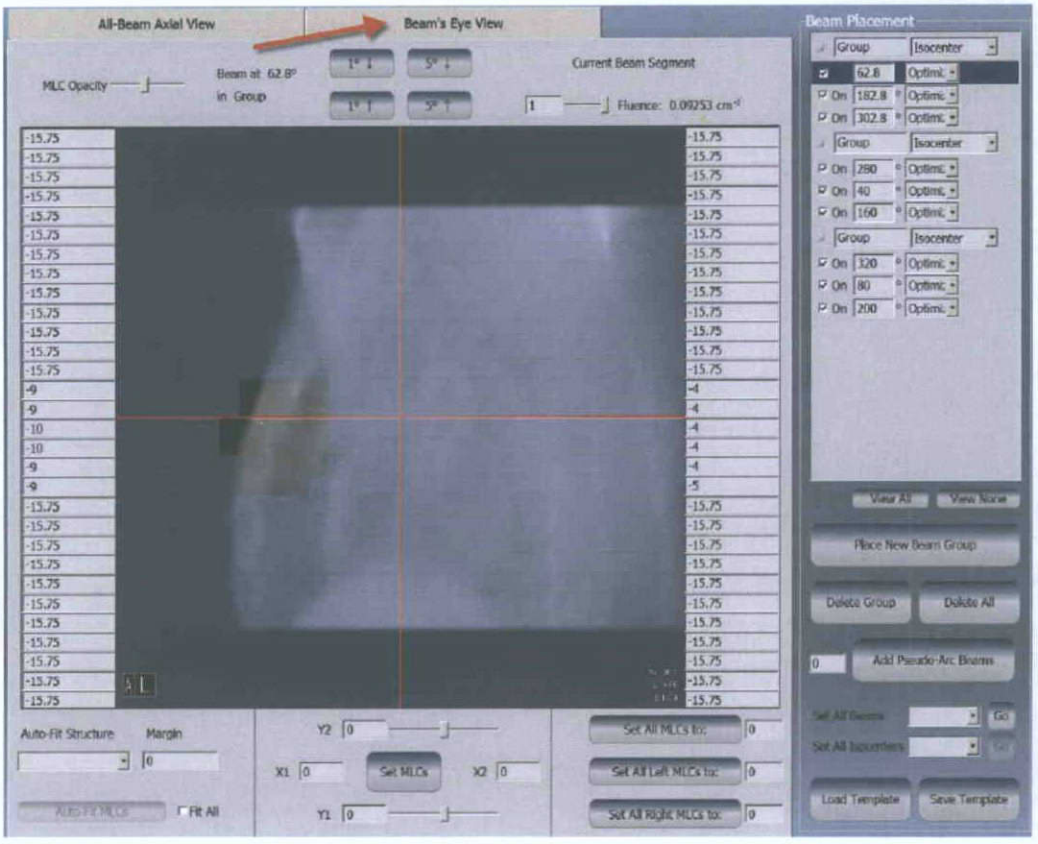
Click > Save Template from the Beam Placement group box.

3.15 Setting MLCs for Conformal Beams

To set the beam shape for conformal beams:

Click > Beam's Eye View at the top of the **Beam** screen

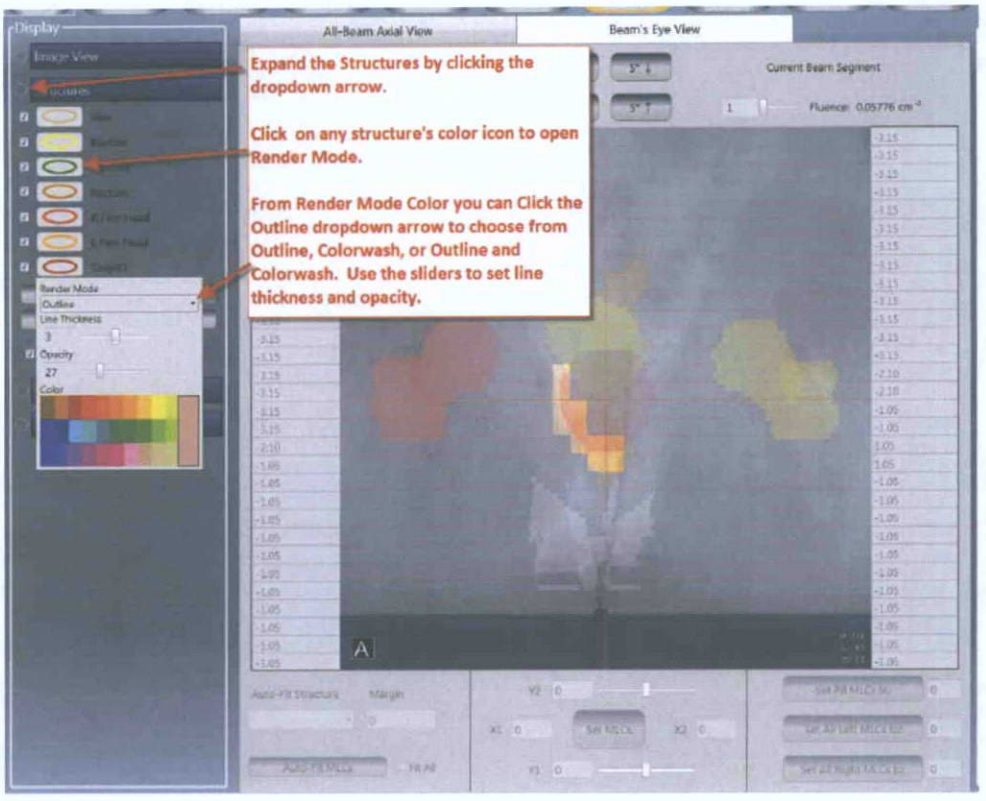
❖ The Multileaf Collimator Setup screen will open.



Set the collimator openings **manually by clicking and dragging** or using the **Auto-Fit Structure** dropdown box to shape the collimator openings to conform to a saved contour template. You can also set the MLCs by **setting centimeter values at the bottom right of the screen.**

NOTE **IMRT beams cannot be manually set.**

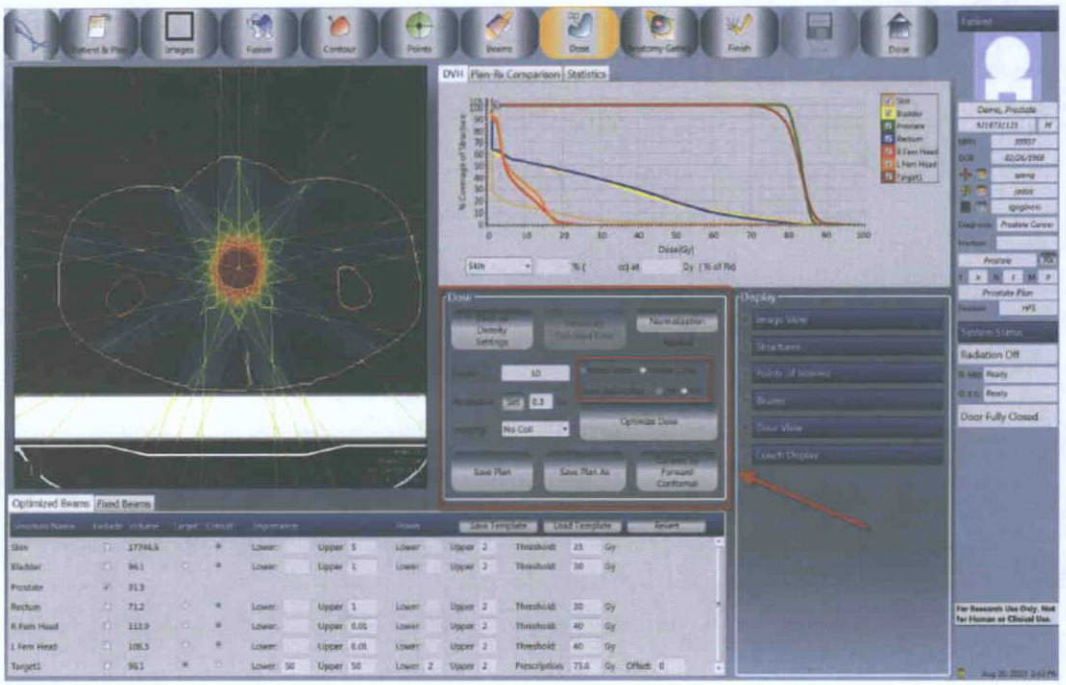
3.16 Viewing Structures from Beam's Eye View



3.17 Dose

The Dose interface uses the inputs of the radiation target's composition, its geometric relationship with the beams and the size and shape of the target and critical structures to calculate the delivery MLC segments and fluence.

You can also choose between Pencil Beam and Monte Carlo algorithms for Dose Optimization.



3.17.1 Specifying Dose Constraints

For IMRT and Optimized Conformal beams you must add dose target prescriptions and critical structure thresholds on the Dose screen. You must also set prescription dose parameters for power, relative upper and lower importance and offset for each structure.

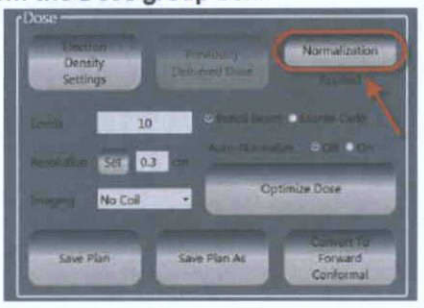
3.18 Introduction to Dose Normalization

After a dose distribution is computed, the distribution can be normalized to a point or a dose volume constraint. Normalization is a scaling of fluences to achieve a specific targeted dose to a particular point or volume.

The normalization scales the dose distribution by a constant factor over the irradiation region. The dose distribution, statistics, prescription comparisons, and DVH are recomputed to maintain the same ratio.

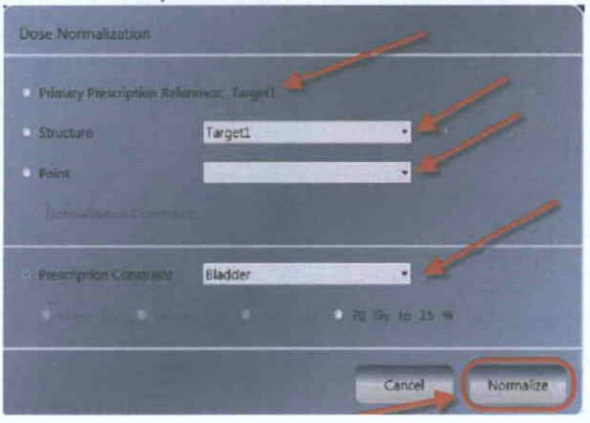
3.18.1 Normalizing the Dose

1. Click > Normalize from the Dose group box.



❖ The Normalize Dose information box will open.

2. Select from the information box the primary prescription, structure, point or prescription constraint you want to use to normalize the dose.



3. Click > Normalize.

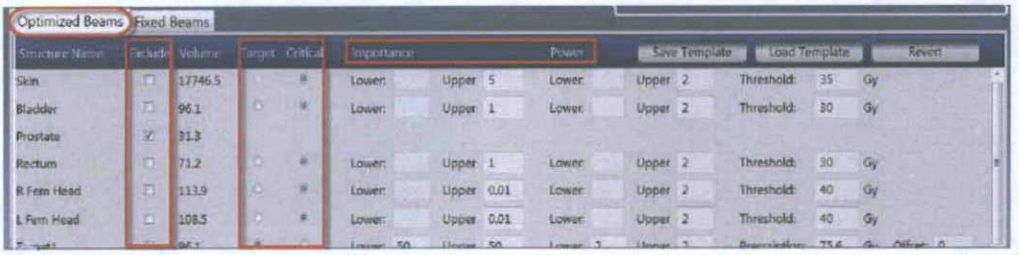
NOTE Whenever lowering the dose levels in a dose distribution results in beam-on fluence levels that drop below the minimum allowed beam-on time, the dose is reoptimized without those beam-on segments.

3.19 Optimizing Dose for IMRT and Conformal Beams

To optimize **IMRT** or **Optimized Conformal** beams you must set the optimization constraints—target(s), critical structures, importance, power, threshold, and prescription/offset.

- ❖ **At least one structure must be identified as a target structure when planning IMRT or Optimized Conformal beams.**

You can **Exclude** structures, which means they will not be included in the optimization algorithm, or you can set structures as **Target** or **Critical**.



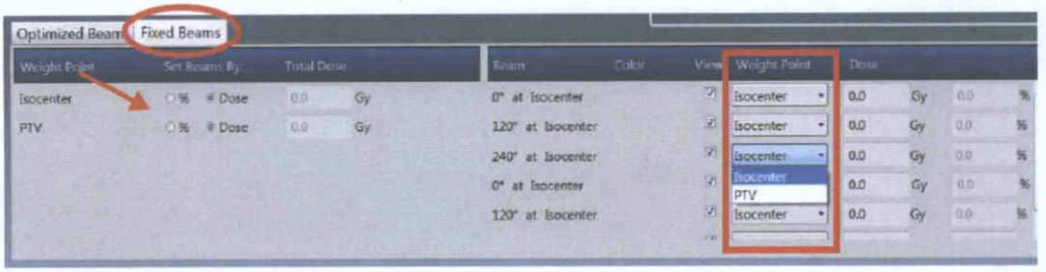
For critical structures, you can set a **radiation dose threshold**. For target structures you assign a **prescription dose**, along with an **offset for fine adjustments**.

You also set the **importance** and **power** for each structure. ***Higher importance and power settings for a structure increases the numerical significance of those structures in the optimization algorithm.***

Electron densities must also be set. See Manually Setting Electron Densities on page 109.

Typically CT images are used to automatically set electron densities because their high-contrast resolution distinguishes between tissues that differ in physical density by less than 1%. You can also manually set the electron densities.

3.19.1 Fixed Conformal Beams



Each fixed conformal beam has one beam-on segment and one fixed shape with uniform fluence.

To optimize the fluence of the single segment of each beam, specify a weight point for each fixed conformal beam based on either the dose or percentage. Each beam will be optimized individually and have its own fluence.

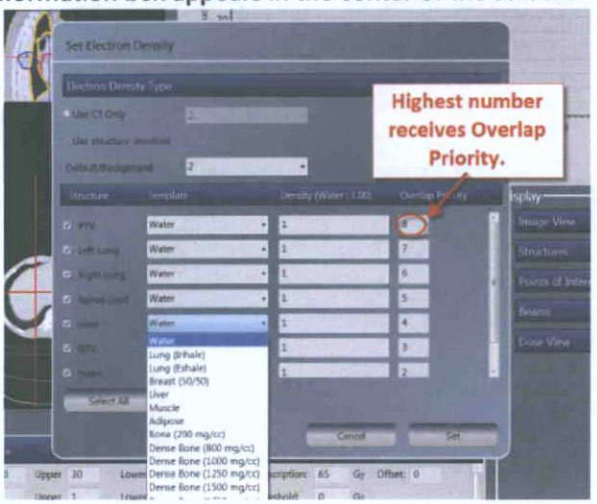
3.19.2 Manually Setting Electron Densities

To optimize the dose for a treatment plan, electron densities must be set. See CT to Electron Density on page 165.

To manually set electron densities use the **Set** button next to the **Specified structure densities** on the **Dose** screen.

Click > Set.

- An information box appears in the center of the screen.



Choose any structure from the information box template and **Click** on it to open the dropdown menu.

You can set densities and priority for each structure from these dropdown menus. For each structure in your plan, select the template electron density for that structure. **Priority** is used if one voxel is within multiple structures. **The structure with the highest priority (highest number) will be used as the electron density setting for the shared voxel. In other words, a setting of 3 receives higher priority than a setting of 1.** Selecting **Custom** allows you to type in an electron density relative to water.

When you are satisfied with your selections,

Click > Set.

After you have set prescription values and defined constraints for tumors and critical structures, you can optimize the dose.

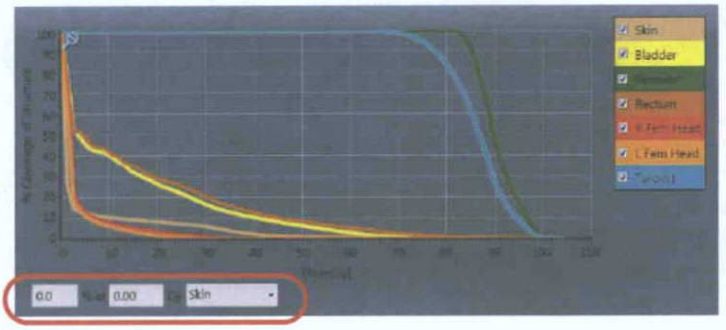
Click > Optimize Dose.

- A progress bar will appear in the center of your screen. The progress bar disappears when the optimization is finished and the **Statistics** box will be populated and the **DVH** will show the dose distributions. The dose is optimized using beam angles, relative intensities, and beam shaping. To view the beams go to the **Beam's Eye View** screen.

1861

3.19.3 DVH Plot Interpretation Tools

The Dose Volume Histogram (DVH) graphically displays the optimized radiation distribution that the patient will receive if the plan is delivered.



The y axis represents the percentage of covered structure and the x axis is the Dose (Gy).

Click the arrow in the plot display at the bottom left of the DVH to select a structure and show the percent of dose in Gy that structure would receive.

You can also **click on any plot line** in the DVH and the display will show the percent Gy that structure would receive.

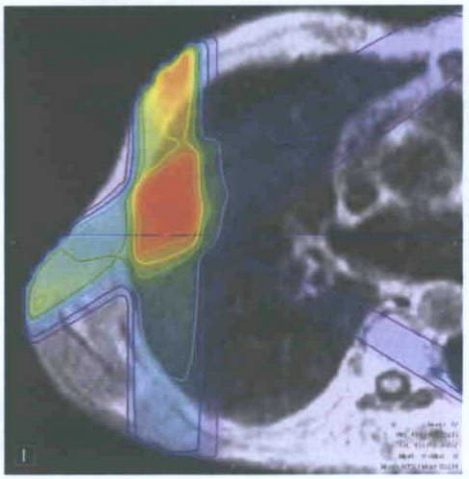
You can zoom by clicking and dragging on the DVH plot. You can reset the zoom by clicking on the magnifying glass in the upper left corner.

You can select which structures you want to view by checking or unchecking the box next to the structure on the structure list , which is on the right of the DVH.

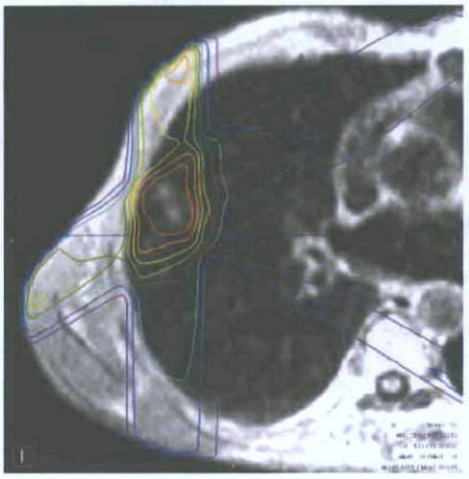
3.19.4 Isodose Lines

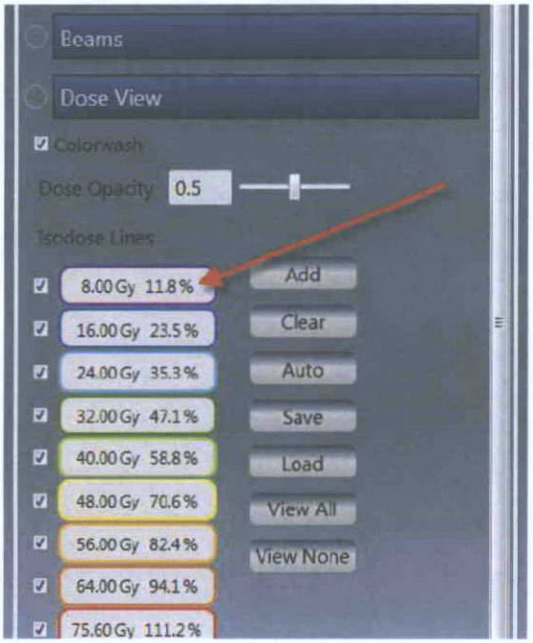
Isodose lines depict the lines of equal dose measurement and can be viewed with or without colorwash.

With Colorwash



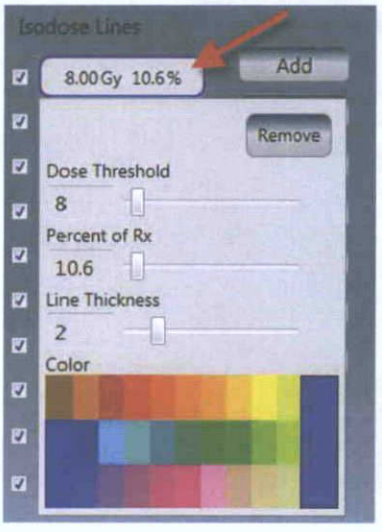
Without Colorwash





To change or adjust the color of any isodose line, **click on any of the isodose icons and the Render Mode** will open underneath the icon.

From **Render Mode** you can set the **Dose Threshold, Percent of Rx, Line Thickness and Color.**



3.20 Calculating the Plan Dose

You can select either **Pencil Beam** or **Monte Carlo** algorithm for dose calculation.

1. **Click** > the option button next to the algorithm you want to use.
2. **Click** > **Calculate and Display Dose** button.

❖ A progress bar will appear in the center of the screen.

3.20.1 Difference Between Monte Carlo and Pencil Beam Algorithms

Monte Carlo dose simulation is an accurate method used to simulate the physics of particle interactions. In this method, ionizing photons with randomly sampled energy levels and directions are considered. As these photons pass through beam openings, the interactions between photons and materials with different geometries, attenuation factors etc. are also considered. Every photon history is initiated by random samplings based on attributes of the radiation source. This photon and its secondary particles are then traced through simulated interactions with surrounding materials and patient and/or phantom volumes. Each photon history may have energy deposited to only very small portions of the patient/phantom volume. Consequently, large numbers of photon histories must be simulated and an average of the deposited energy must also be taken to get a stable dose profile at each point. When conducted carefully and thoroughly, this physics-based method is very accurate for dose simulation.

Pencil Beam calculation, unlike Monte Carlo, is an analytic process in which one splits beam openings into small blocks. Each block is called a beamlet which has a predefined dose profile. In pencil beam dose calculation, this predefined profile is scaled for each beamlet and the summation of the scaled profiles yields the overall dose. The predefined dose profile for pencil beam comes from fittings based on actual dose measurements, and is therefore considered an accurate analytical method for dose calculation.

3.21 Dose Computation

The ViewRay System™ employs two methods of dose computation: Pencil Beam and Monte Carlo.

The Pencil Beam computation is a finite-sized pencil beam model with full divergence and a depth-dependent penumbra. It uses density scaling by ray tracing from the source to every voxel in the beam. It is a very efficient dose computation algorithm and is used in the beamlet model of the dose optimization algorithm for IMRT beams. The Pencil Beam model is most accurate when computing dose to a homogeneous water media. It matches the Monte Carlo model very precisely in water, but the Pencil Beam does not consider the effects of the MRI magnetic field on the dose distribution.

The Monte Carlo algorithm performs a simulation of the photons that are created in the source volume and the interactions of those particles with the source capsule, transports that calculation to the patient and considers the MRI's magnetic field influence on the secondary electrons. The Monte Carlo algorithm is more accurate than the Pencil Beam model in heterogeneous media and in interfaces to low density media where the MRI magnetic field can significantly influence the dose distribution. However, the Monte Carlo algorithm is not efficient enough to compute beamlets for the dose optimization.

3.22 Dose Optimization

The ViewRay System™ can optimize the dose from intensity modulated radiation therapy (IMRT) beams or Conformal Optimized beams. Dose optimization uses a mathematical model of large-scale optimization in order to determine the beam fluences that are required to produce a conformal dose distribution. The quality of the coverage of targets and the sparing of critical structures is controlled by setting dose optimization parameters for each target and critical structure in the model.

Beams are selected by the user and can be defined as IMRT or optimized conformal beams and placed at the desired gantry angles and isocenters for optimization.

Beams defined as Conformal will contribute dose to the treatment plan according to their normalization and will be considered in the optimization model. Also, the optimization model will consider any existing dose that has been imported into the plan or recorded by the system. Pre-existing dose or dose from conformal beams is not modified by the dose optimization algorithm. IMRT beams are discretized for each beam into 1.05 cm by 1.05 cm beamlets and the dose to the patient's targets and critical structures for each beamlet is considered in the optimization.

The optimized dose distribution is controlled by setting dose optimization parameters for each critical structure or target in the plan. Critical structures and targets can also be excluded from the optimization model. Excluded critical structures and targets will default to being included in the underlying structures or targets included in the model. Each segmented anatomy in the plan should be identified as "Target" or "Critical" using the option buttons in the "Optimized Beams" tab.

The dose optimization model requires at least one target structure and the skin structure for the general sparing of tissue to operate. The "Skin" critical structure is not optional and must exist for the dose optimization feature to function. The planning system does not alter beam aspects set by the user. An efficient algorithm is then employed to solve the model. Once solved, the fluences are discretized to levels selected by the user and a leaf-sequencing model is solved to produce beam segments that efficiently delivery the fluences while producing a minimal tongue and groove effect.

The dose optimization model is a convex nonlinear programming model where all voxels have a penalty function associated with them according to their underlying structure or target. In this model, undesirable dose distributions incur a penalty where the more undesirable they are the higher the penalty becomes. In the optimization model each voxel is associated with a single target or critical structure.

Note that this is only true for the optimization of the model and not true for DVH information computed by the system. If multiple targets and critical structures overlap a given voxel, a priority list is used to determine which critical structure or target dominates. Typically, targets dominate critical structures as they are intended to be treated to a prescribed dose, and structures with more serious side effects dominate less important critical structures.

The priority of the targets and critical structures can be set in the Optimize Beams tab. The dose optimization parameters for the penalty functions are set separately for each structure or target in the Optimized Beams. Each target has 6 parameters and each critical structure has 3 parameters. The penalty function for targets is designed to try to make all of the dose in the target have the same value.

A prescription dose, D_{Rx}, gives the desired dose for target coverage. An offset can be used to move the dose distribution to a higher dose level by placing the minimum penalty at a higher dose. This is typically required if more than 50% coverage of the target is required, and penalties for voxels with doses lower or higher than the prescription dose plus the offset are specified. There are two types of penalty parameters for each case: a) an importance or scaling factor and, b) a power for exponentiating the dose difference from the prescription dose plus the offset.

For voxels with doses lower than the prescription dose plus the offset, the "Lower Importance" (LI) is the scaling factor, while for voxels with doses higher than the prescription dose plus the offset, the "Upper Importance" (UI) is the scaling factor.

For voxels with doses lower than the prescription dose plus the offset the "Lower Power" (LP) is the exponentiating factor, while voxels with doses higher than the prescription dose plus the offset the "Upper Power" (UP) is the exponentiating factor. Critical structures are treated like targets, except the prescription dose is changed to a threshold dose (T_{Rx}), so there is not an offset, and they do not have lower penalties. Please refer to figures 1 and 2 below.

The user sets each value of the dose optimization parameters to values that correspond to the objectives in the patient's prescription and then solves the model by pressing the Optimize Dose button. Note that the prescription dose volume and point constraints set in the Prescription Tool are not directly tied to the dose optimization parameters. Once the model is solved, the prescription criteria can be reviewed in the Plan-RX Comparison tab. If the plan is found to not satisfy the prescription, the parameters are adjusted in a logical manner and the model can be reoptimized by pressing the Optimize Dose button.

Typically, one increases an importance to make a critical structure or target penalty less important to the model. The exponentiating power is increased if one wants to make larger deviations much more severe than lesser ones. The thresholds and prescription doses are typically set to their clinically expected values. The offset is used to move the target dose distribution coverage of a given percentage to a higher value. The system allows a structure to be excluded from the optimization.

NOTE: If a structure or target is excluded, the optimizer will not consider it during optimization; however, if the underlying voxels are shared then ray tracing is applied to them during dose calculation, but, as stated previously, they will be excluded from dose optimization. If a structure or target is excluded, the optimizer will not consider it during optimization; however, if there are underlying shared voxels then ray tracing is applied to those shared voxels during dose calculation, but, as previously stated, the voxels are excluded from dose optimization.

Figure 1: Critical Structure

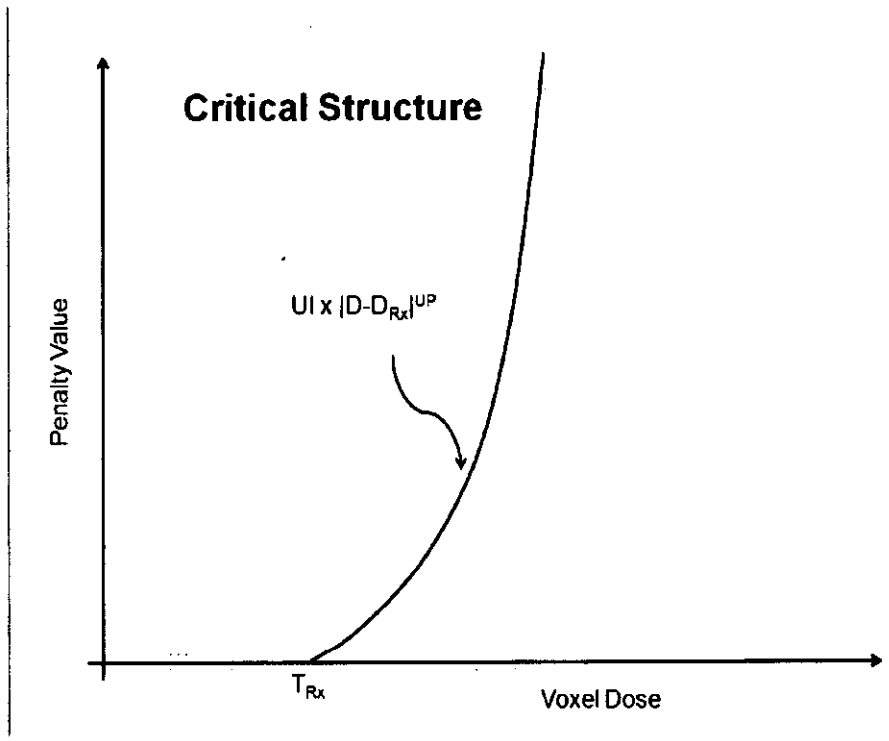
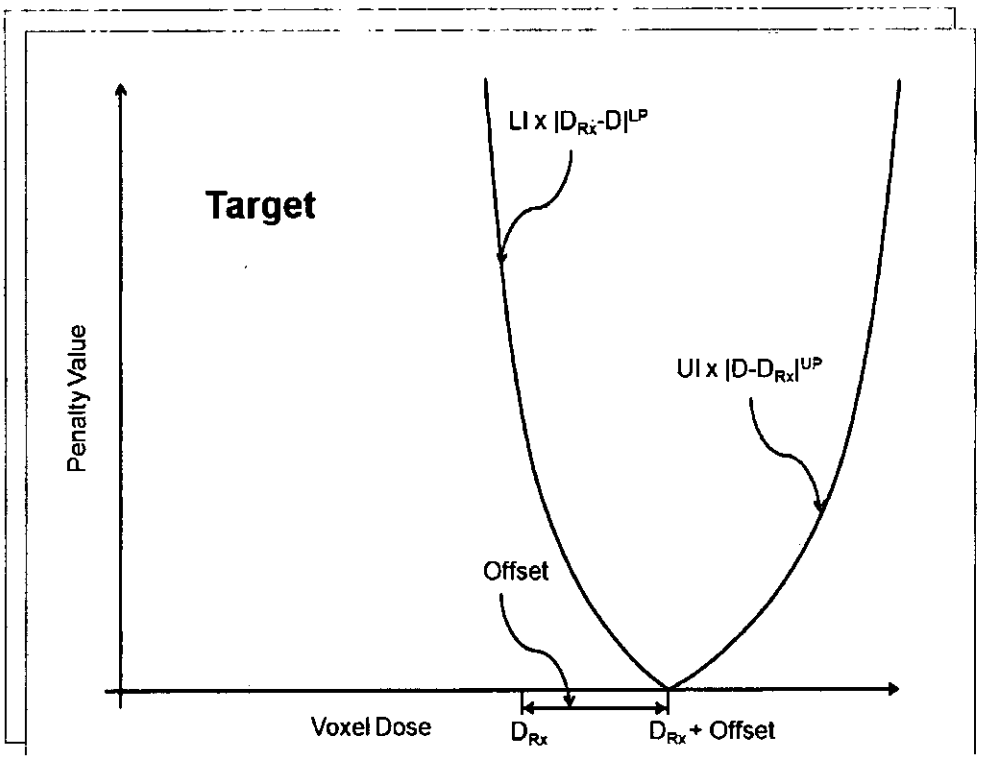


Figure 2: Target

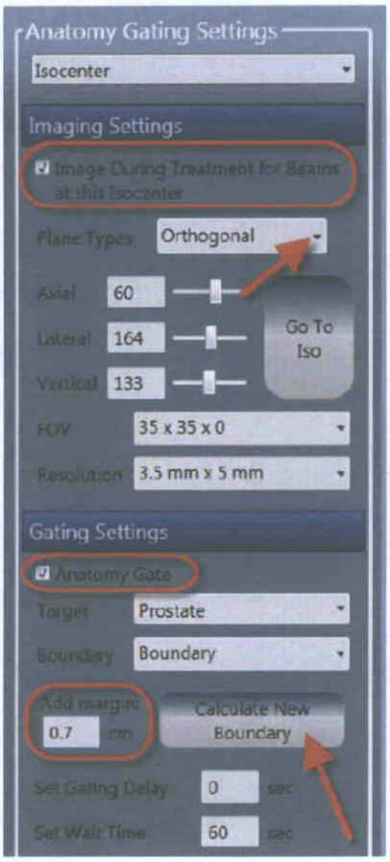


3.23 Anatomy Gating

Anatomy gating is an optional feature that allows you to set up spatial and time tolerances for pausing a treatment delivery. You can stop treatment when a structure moves too far outside of a pre-set boundary and you can determine how long you want treatment delayed.

The spatial tolerances for anatomy gating are set by selecting a target structure and a boundary structure. Whenever the target structure goes beyond the boundary you set, then treatment is paused until the target moves back into range.

The target structure must be a pre-existing structure, imported or created on the structures screen.



Use the **Imaging Settings** group box to set up to the gating image planes.

Use the dropdown arrow next to **Plane Types** to select either **Axial, Lateral, Vertical** or **Orthogonal.**

Use the sliders to move through the slices to select the images you want to use for gating or use the **Go to Iso** button to move all the slices to Isocenter.

You set the **Target** and **Boundary** structures from the dropdown boxes in the **Gating Settings** group box. You can also select a structure and then add a margin (in cm).

3.23.1 Enabling Anatomy Gating for a Plan

From the **Anatomy Gating Settings** group box located at the right of the image window **Check > Perform Gating for This Plan.**

3.23.2 Real-Time Anatomy Gating

During delivery, the target structure is auto-contoured in real time. The boundary structure stays constant. If any part of the target structure moves outside the boundary, then the radiation beam is turned off. When the target structure moves completely back inside the boundary, then the beam is turned back on and treatment continues.

3.23.3 Anatomy Gating Delay

Gating Delay can be set so the system will gate off only after the target has moved outside the boundary for the pre-set delay time period.

Wait time can be set as well. If the target has stayed outside the boundary for longer than the wait time, the system will terminate delivery and allow the user to redo the patient setup.

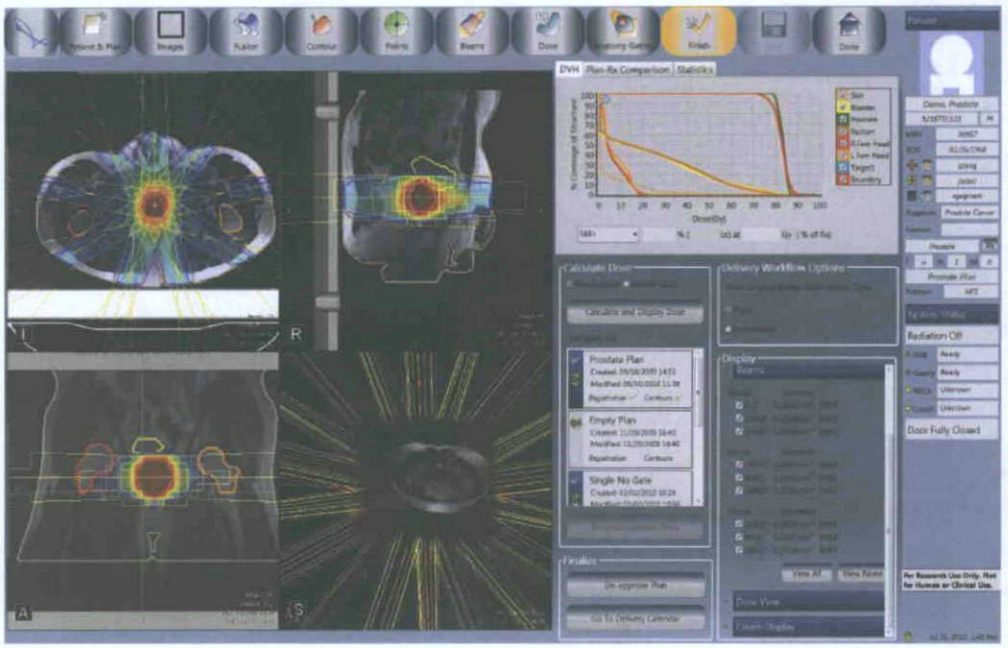
3.24 Tracking Points

See Tracking Points on page 147

3.25 Treatment Plan Evaluation and Approval

From the **Finish** interface you can compare the plan to the prescription, view the DVH and plan statistics, and choose either Pencil Beam or Monte Carlo algorithms for dose calculation.

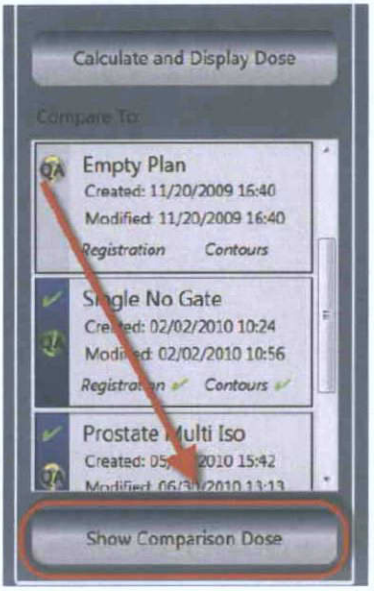
You can display the Dose, compare two plans, approve a plan, save the plan, export the plan to DICOM, or go to the **Delivery Calendar** and treat immediately with an approved plan.



3.25.1 Comparing Two Different RT Plans

To compare a currently loaded plan with another plan:

1. Click the dropdown arrow to open the **Compare To** dropdown box.
2. Click on the plan you want to use for comparison from the dropdown box.
3. Click > **Show Comparison Dose**



❖ A progress bar will appear in the center of the screen. When the comparison is finished the progress bar will disappear and a new screen will open with the current plan and the comparison plan side by side.

The current plan images are on the left and the comparison plan images on the right.

The top **Display** group box is for the original plan images on the left and the bottom **Display** group box is for the comparison plan images on the right.

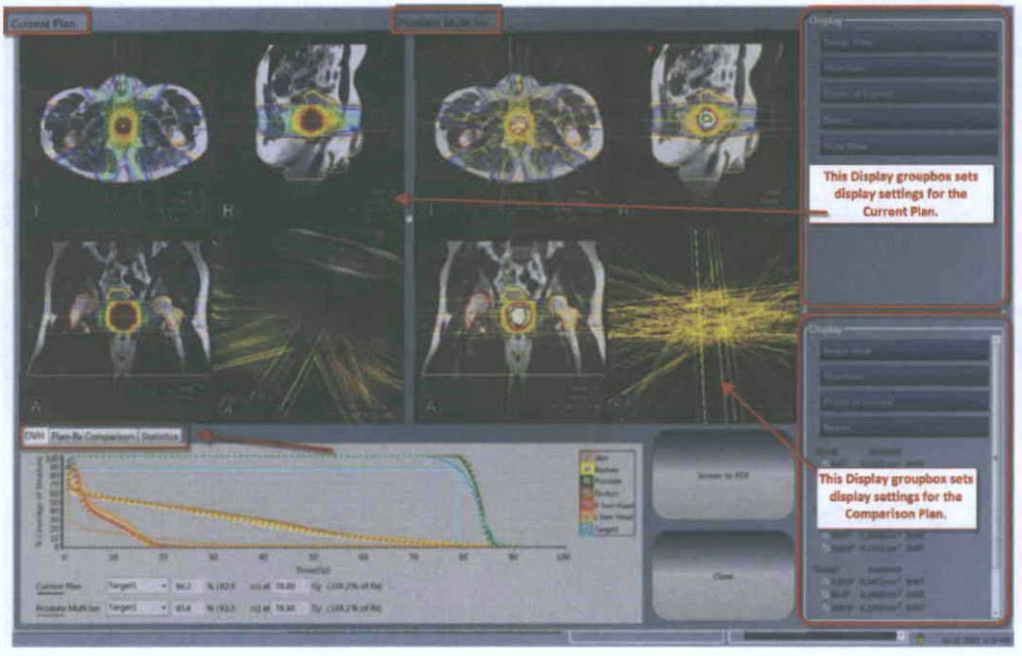
There are three comparison modes—**DVH, Plan-Rx Comparison**, and **Statistics**. I

In the **DVH**, the solid plot lines are the current plan dose in Gy and the dashed plot lines are the comparison plan dose in Gy to structures.

Plan-Rx Comparison compares the prescription constraints to both the current and the comparison plans.

Statistics displays—side by side—the **min, mean** and **max** for each contoured structure for both plans.

Click on the menu tabs to move back and forth between the various comparison formats.



3.25.2 Approving a Treatment Plan

The **Approve Plan** and **Go To Delivery Calendar** command buttons are at the bottom center of the **Finish** interface.

To approve a plan:

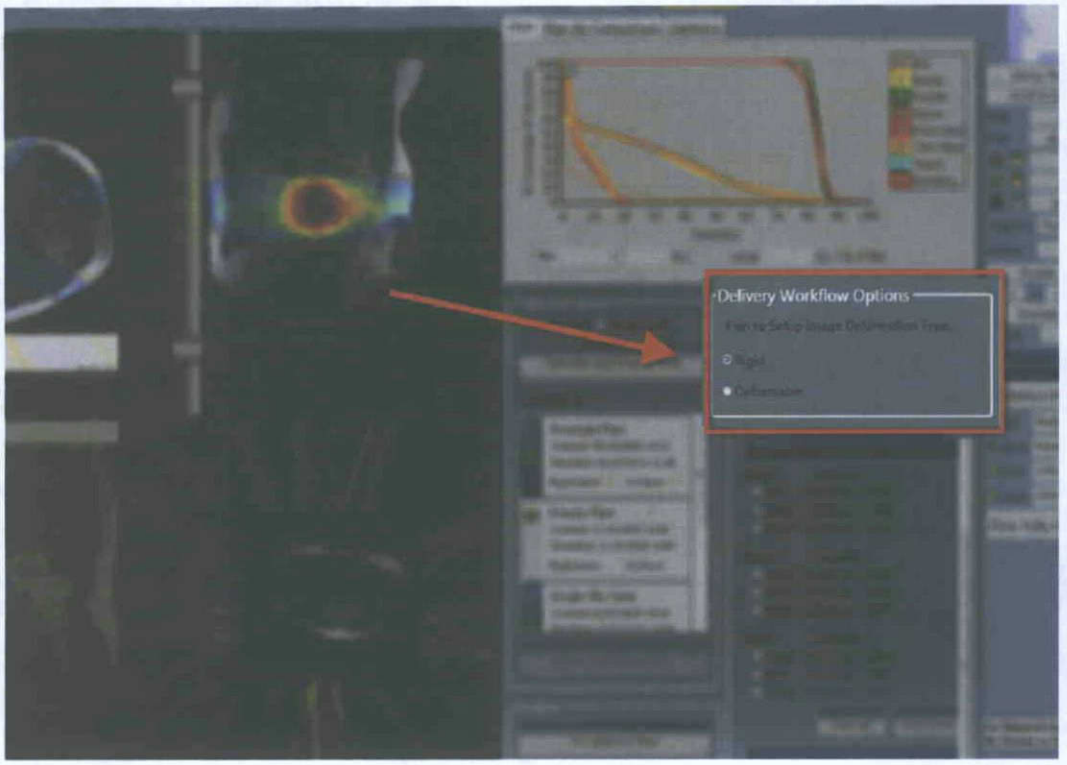
1. Click > Approve Plan
   - ❖ A progress bar will open in the center of the screen.

After the plan has been saved and approved, the command buttons will update. The buttons will now be **Un-approve Plan** and **Go To Delivery Calendar.**

NOTE **ONLY APPROVED PLANS CAN BE DELIVERED.**

3.26 Setting Delivery Workflow Options

You can set the registration mode for the **Treatment Delivery** image registration to rigid or deformable on the **Finish** screen.



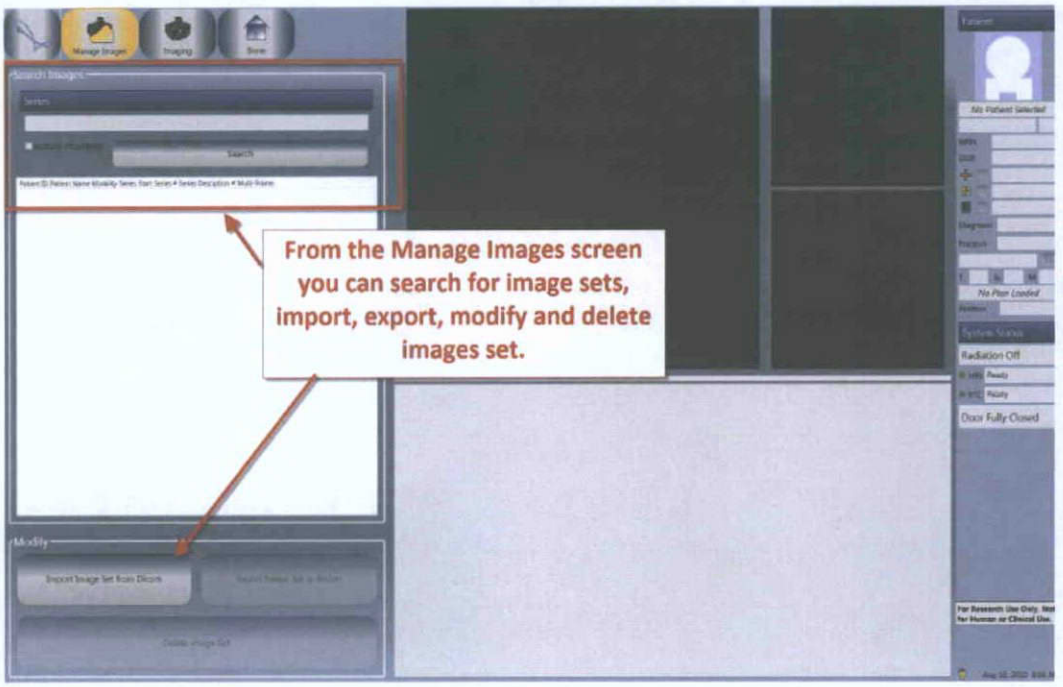
[redacted]

THIS PAGE INTENTIONALLY LEFT BLANK

1578

Chapter 4: Imaging

The **Image** interfaces are available from both the **Treatment Planning** and **Treatment Delivery** workflows. You can use the interfaces to search for images, import and export and delete images, as well as acquire new MR images of your patients.

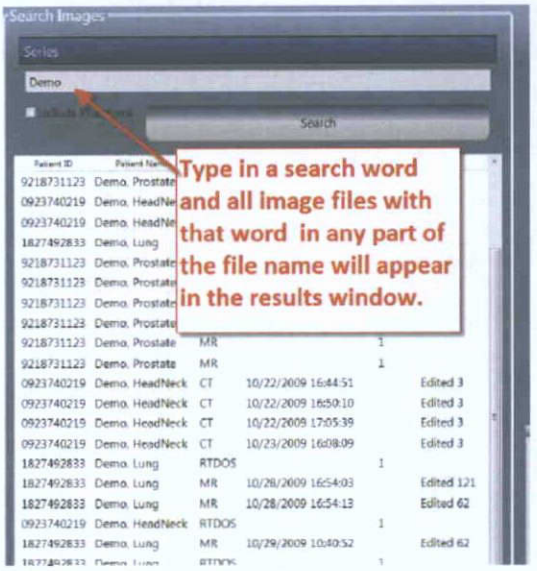


4.1 Searching for Images

To search for images:

1. Type a search word.
2. Click > Search.

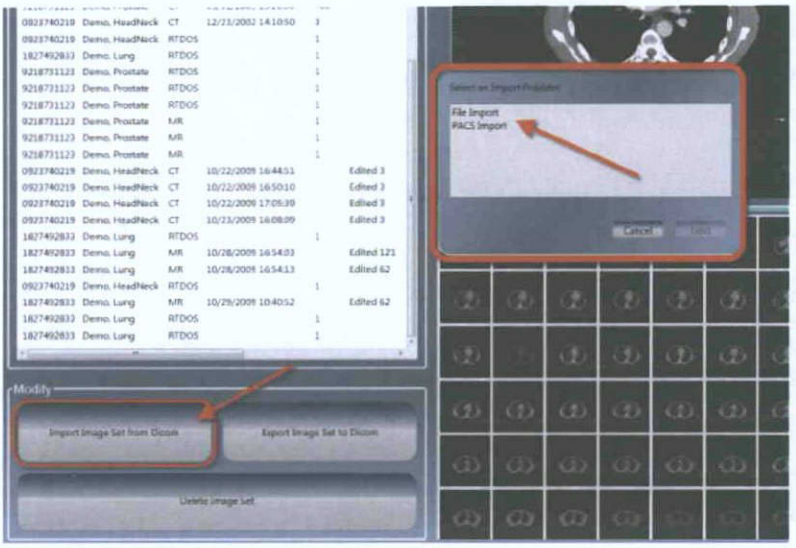
❖ The search results will appear in the results window.



4.2 Importing Images from PACS

To import images:

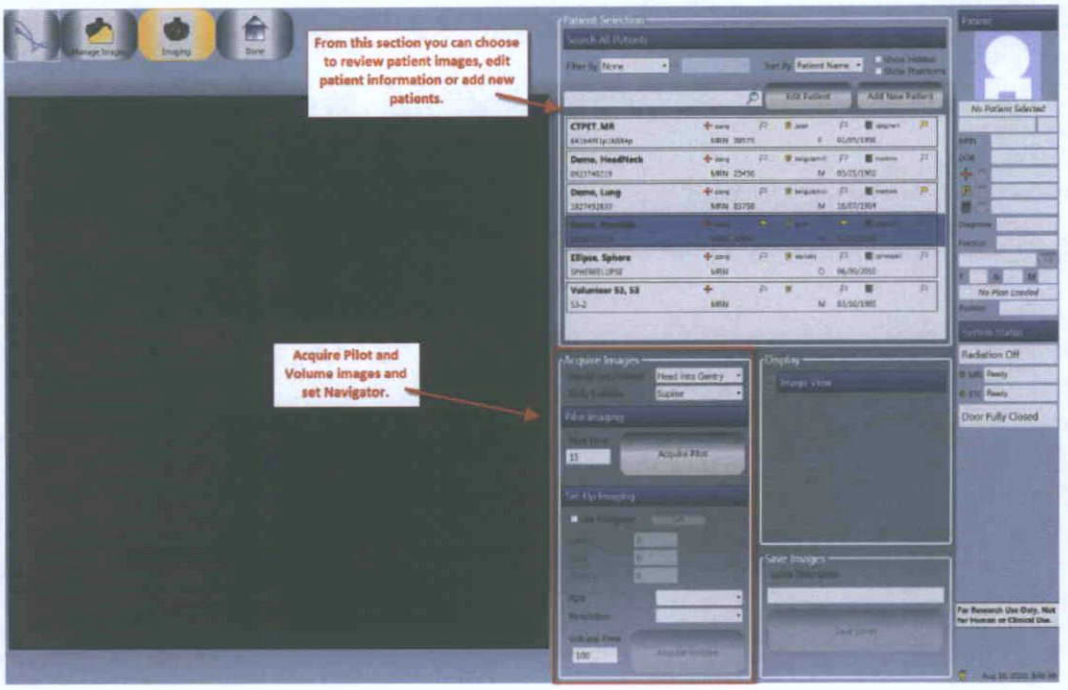
1. **Click > Import Image Set from Dicom.**
   - ❖ **The Select an Import Provider** information box opens.
2. **Click > PACS Import.**



- ❖ **The Choose your destination** information box will open.

4.3 Imaging

From this interface you can search for patient images, edit patient data, add new patients, acquire pilot and volume images and save images.



Chapter 5: Treatment Delivery

5.1 Introduction

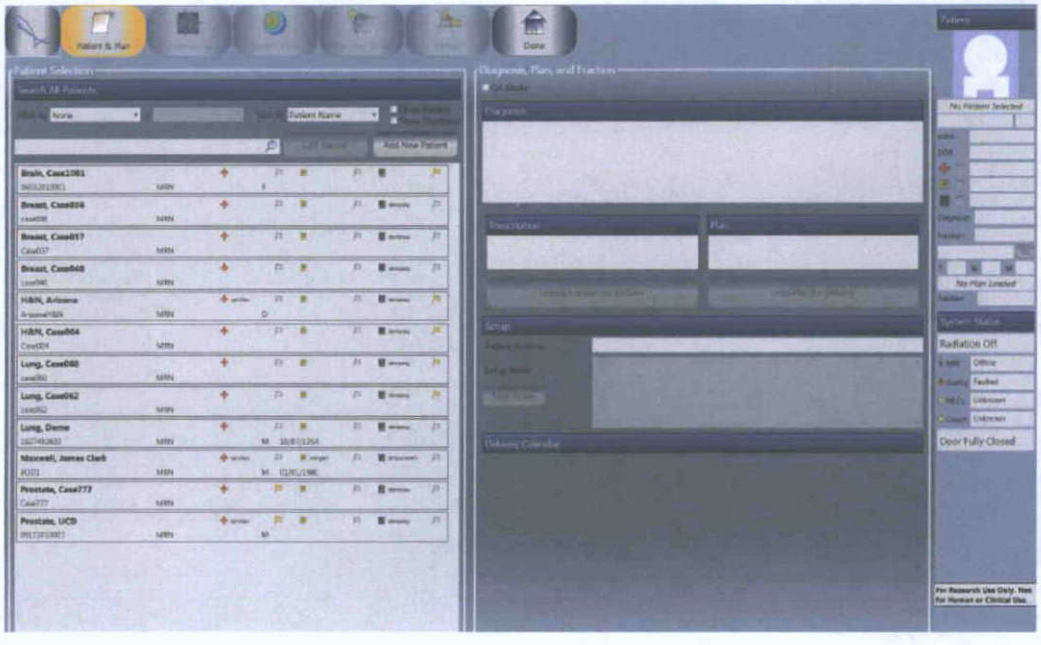
The ViewRay System™ can deliver radiation therapy with fixed conformal, conformal and intensity modulated radiation beams. The beams in an intensity modulated radiation therapy plan are delivered in a "step-and-shoot" segmental mode.

Every treatment plan can be verified in a plan QA available with the system.

5.2 Retrieving the Patient

1. Click > Treat from the menu bar on the Treatment Delivery Home Screen.

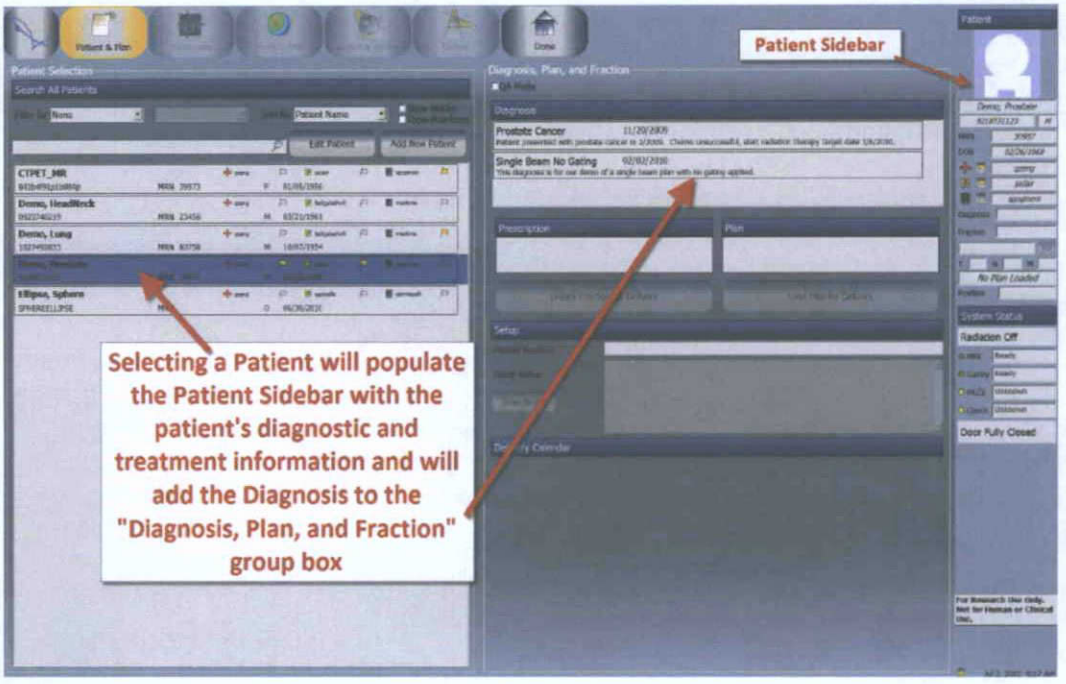
❖ The **Patient & Plan** screen opens.



2. Select the patient you want to treat by **Clicking on the patient** from the patient list in the **Patient Selection** group box.

❖ The selected patient's diagnosis or diagnoses will appear in the **Diagnosis, Plan, and Fraction** group box

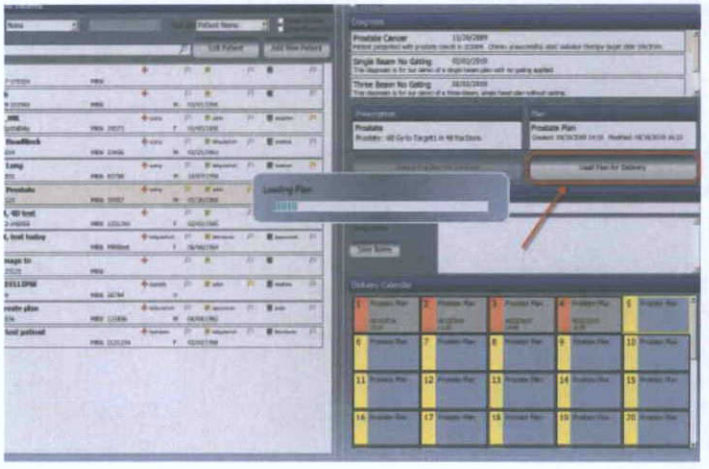
CAUTION: Be careful to check and ensure that the Patient identified in the Patient Sidebar is the patient you are preparing to treat. The ViewRay System may be contraindiated for patients with implants due to the presence of the MRI magnetic field. Patients must be screened for MRI compatibility PRIOR TO ENTERING THE TREATMENT AREA.



5.3 Loading the Plan for Delivery

After retrieving the patient you can load the patient's plan for delivery.

1. **Click** > the **Diagnosis** associated with the Plan you want to load.
   - ❖ A listing for the patient's **Prescription, Plan** and **Delivery Calendar** will appear in the group boxes.
2. **Click** > **Load Plan for Delivery.**



❖ A progress bar will appear.

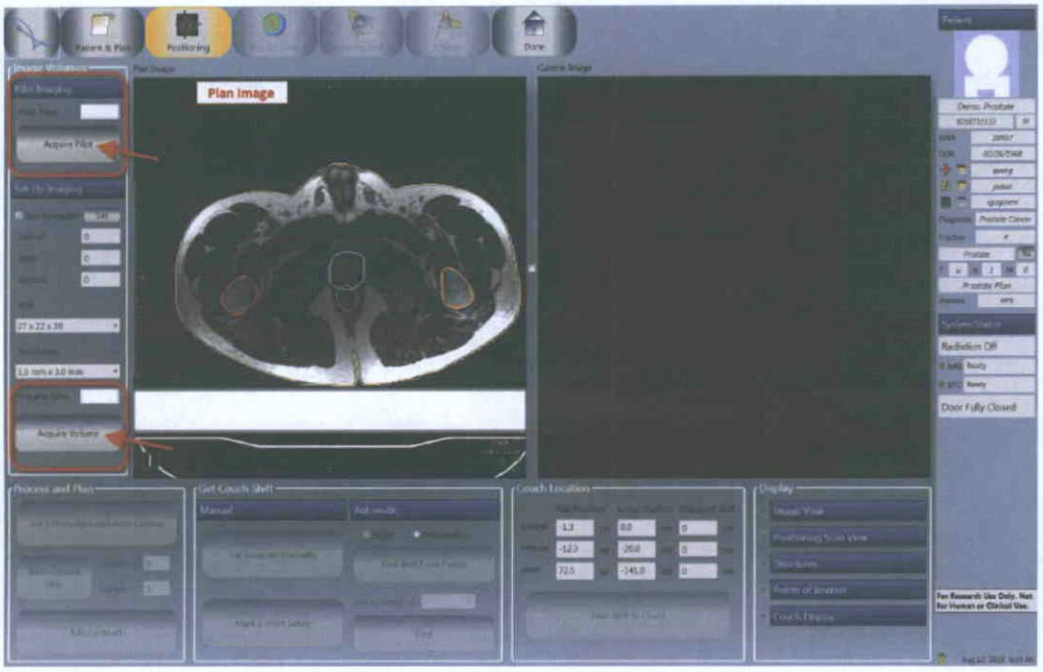
When the progress bar disappears, the plan will be loaded and the **Positioning** screen will be enabled.

You can now **Click the Positioning** tab on the **menu bar** to go to the **Positioning** screen to position your patient for treatment.

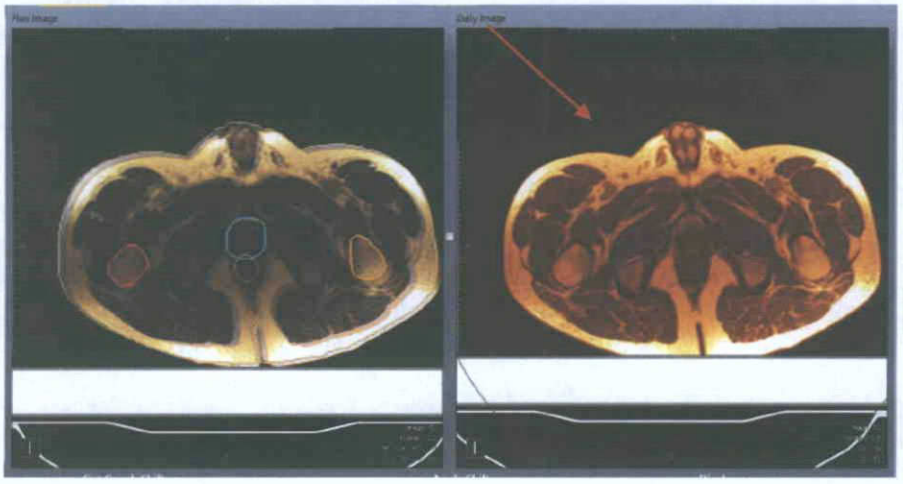
5.4 Using Daily Images to Position the Patient

When you enter the **Positioning** screen the patient's plan image is on the left side. The **Acquire Pilot, Acquire Volume** and **Send Shift to Couch** buttons are enabled.

You can acquire a 15-second, low-resolution, fixed-FOV, gross positioning scan by selecting the **Acquire Pilot** imaging option.



The **Pilot Image** will appear on the right.



5.4.1 Acquire Volume Imaging Option

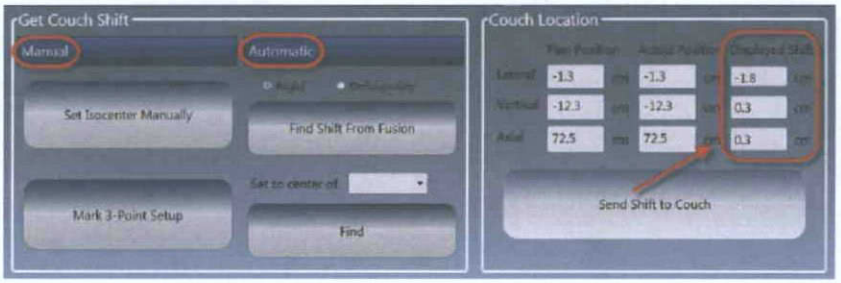
A high-resolution single volume can be acquired in less than 3 minutes using the Acquire Volume feature.

5.4.2 Finding the Couch Shift

You can manually or automatically find the couch shift.

5.4.2.1 Manually Finding the Couch Shift

1. **Click > Set Isocenter Manually** and then **Click** the position on the **Daily Image** where you want the isocenter placed.
   - ❖ The displayed shift is based on the shift at isocenter and will appear in the **Couch Location** information box.
2. **Click > Send Shift to Couch.**
   - ❖ The couch shift will be applied to the **Daily Image** and sent to the **Treatment** and **Control Panels.**



NOTE To move the couch you must use the **Control Panel** on either side of the patient couch or the **Control Panel** next to the console.

You can **Mark 3-Point Setup** to set the isocenter manually and find the suggested couch shift.

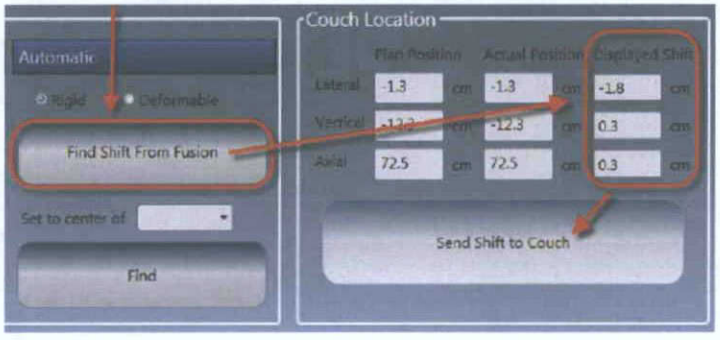
1. **Click > Mark 3-Point Setup.**
2. **Click on a location on the Daily Image** where you want to begin a line, **release the mouse and move it to where you want the line to end and click.**
   - ❖ A dashed red line is drawn on the image.
3. **Move the mouse along the line**—a dashed red line will move with the mouse.
4. **Click at the point away from the first line that you want to select and use to connect with the first line.**

5.4.2.2 Automatically Finding the Couch Shift

To automatically find the couch shift needed to correctly position your patient for treatment you must first **Get Deformation and Auto-Contour the Daily Image** with the **Plan Image.**

1. Make sure you have a **Daily Image.**
2. **Click > Get Deformation and Auto-Contour.**
   - ❖ A progress bar will open.

   - ❖ When the progress bar closes, the deformation and auto-contour is complete and the **Find Shift From Fusion** button will be enabled.
3. **Click > Find Shift From Fusion**



   - ❖ The suggested lateral, vertical and axial shifts will load in the **Displayed Shift** column.
4. **Click > Send Shift to Couch.**
   - ❖ The couch shift is sent to the Couch and Treatment control panels.

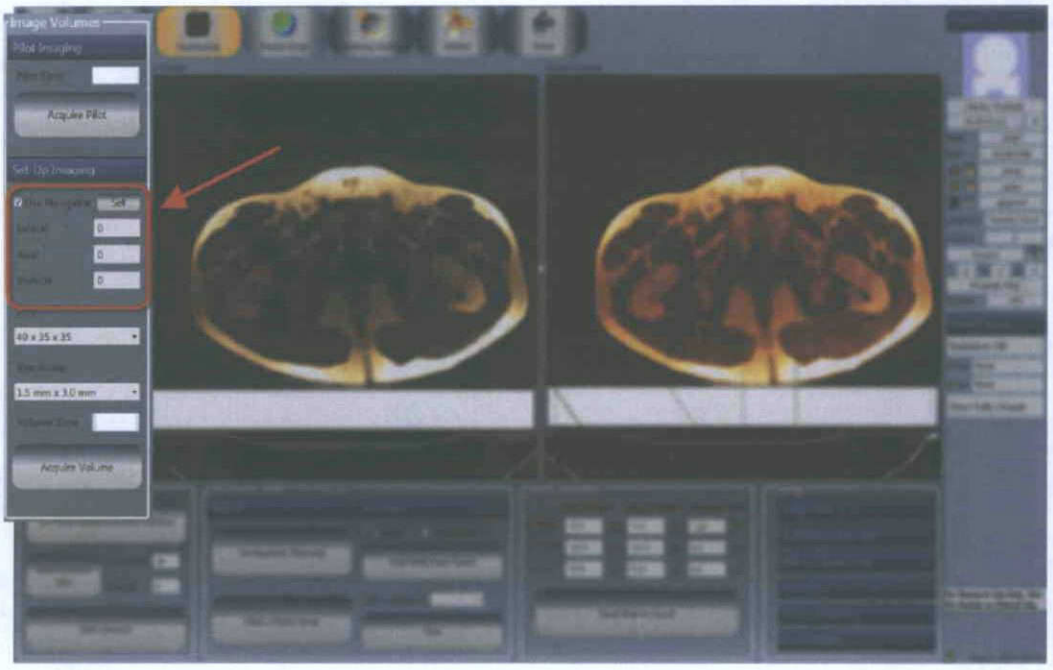
5.5 Using the Navigator

Using the **Navigator** when you are acquiring a volume image of your patient minimizes respiration-induced image ghosting, blurring, and other image distortions which are referred to as "image artifacts."

The 1D navigator uses excitation of a narrow region of interest, such as between the liver and the diaphragm to provide respiratory cycle data.

The Navigator uses this data to trigger image-data acquisition ONLY during the same portion of the respiratory cycle.

Using the navigator function creates a sort of "freeze frame" data acquisition so that each time an image is acquired the patient's anatomy is in the same position.



To use the Navigator:

1. Acquire a Pilot Image.
2. **Check** > **Use Navigator** in the **Set-Up Imaging** group box.
3. Scroll through the Pilot image and locate the apex of the liver.
4. **Click** > **Set** button.
5. **Click** on the **Pilot** image at the place where you located the apex of the liver.

5.6 Predicting Dose

You can predict the plan dose based on the daily images

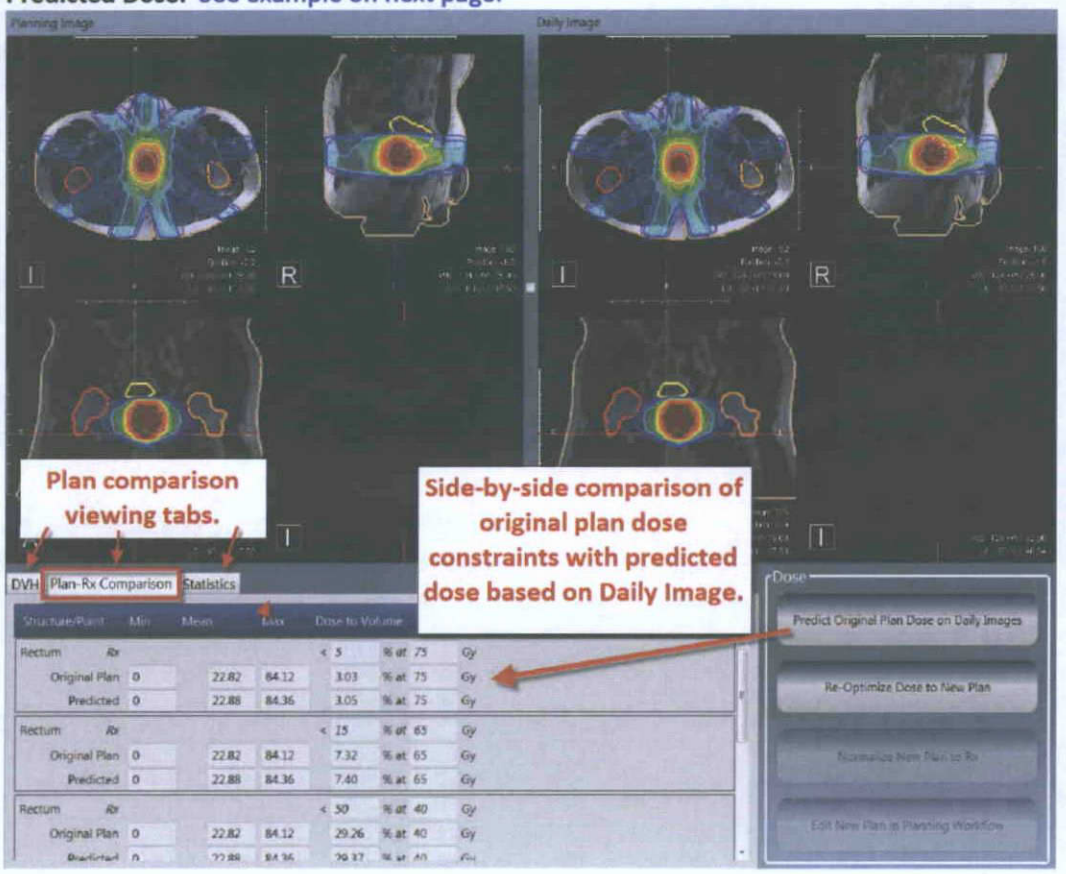
1. **Click > Predict Original Plan Dose on Daily Images.**
   - ❖ The dose prediction is displayed as isodose lines on the **Daily Image** (right side of screen). The original plan dose is on the left.

Use the viewing tabs to see additional comparisons of **Plan** and **Predicted Dose.**

Click on the DVH tab to see the DVH comparison view. The **solid lines** on the **DVH** depict the **Plan dose** and **dashed lines** depict the predicted dose based on the **Daily Image.** See example on next page.

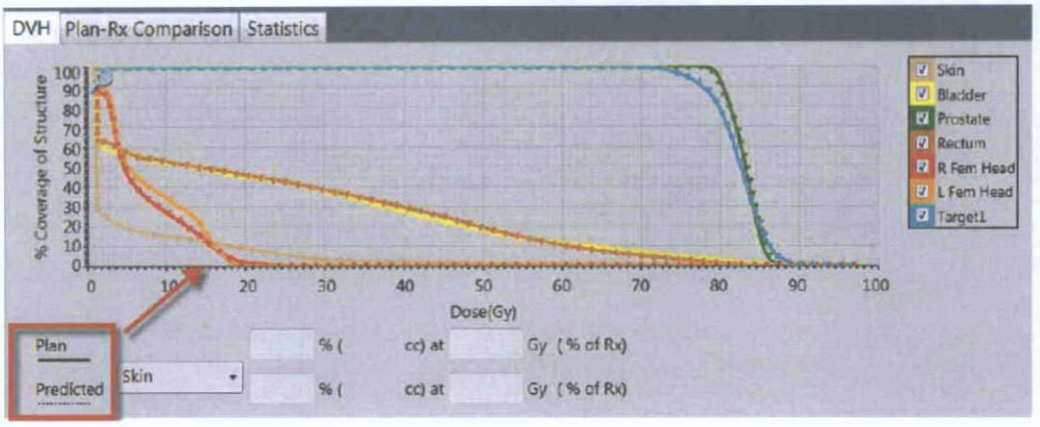
The **Plan-Rx Comparison** shows a side-by-side comparison of the pre-defined tolerances set in the prescription and the original plan and predicted dose based on the daily image. See example below and on next page.

Statistics shows a structure-by-structure comparison of the **Mean Min** and **Max** for **Plan** and **Predicted Dose.** See example on next page.

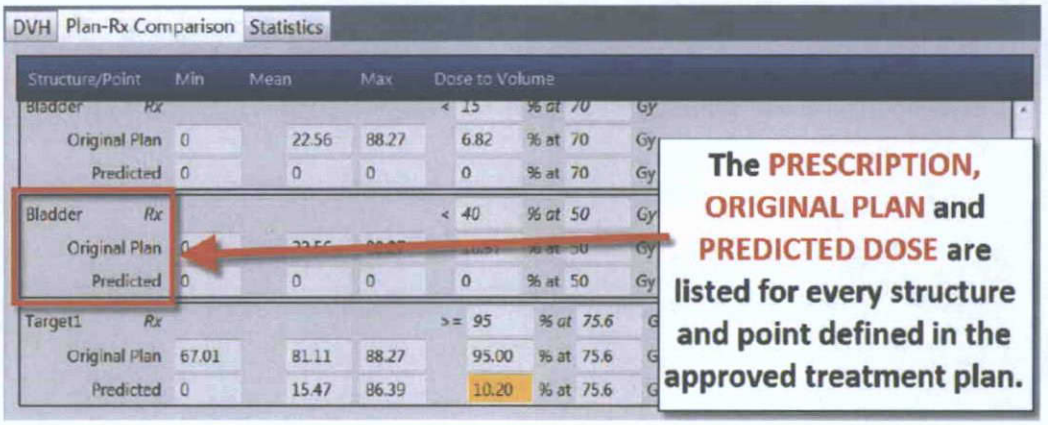


1890

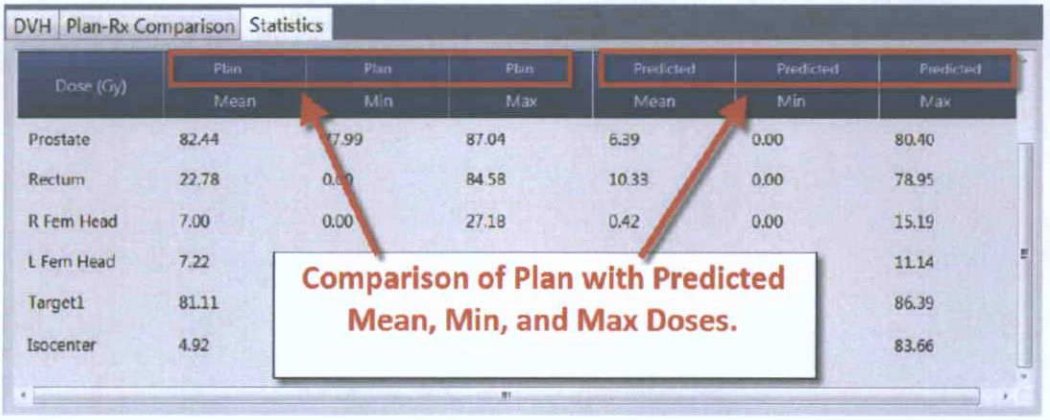
Example of comparison DVH



Example of Plan-Rx Comparison

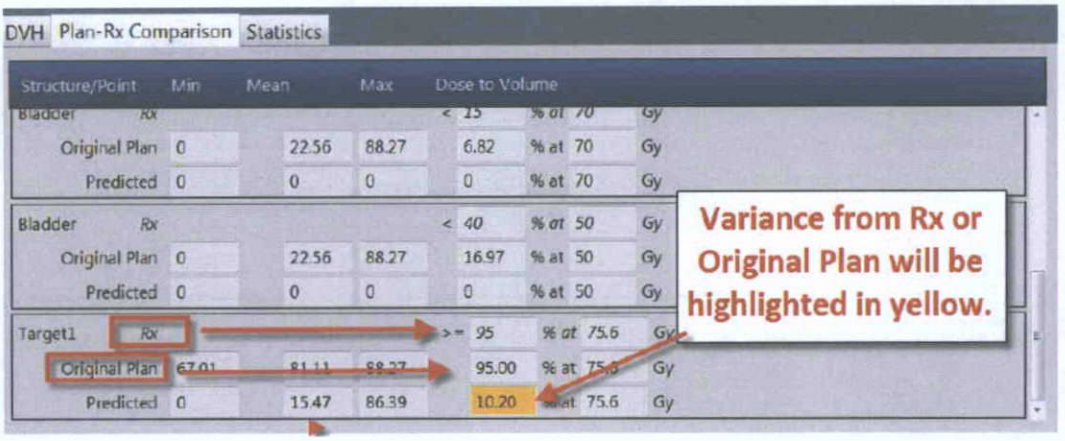


Example of comparison Statistics



5.7 Re-Optimizing Dose to New Plan

After you **Predict Original Plan Dose on Daily Images** you can view the comparison of the original plan and the predicted dose. If there is a variance from the original plan constraints in the **Predicted Dose to Volume**, that variance will be highlighted in yellow on the **Plan Rx Comparison.**



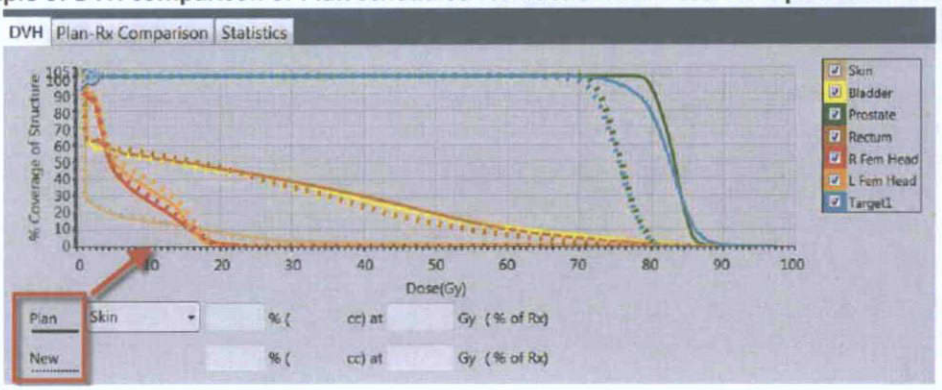
Re-Optimize Dose to New Plan recalculates—based on the **Daily Image** of your patient—beam angles, MLC shapes and fluence and produces a new plan to meet the prescription constraints.

To begin a re-optimization:

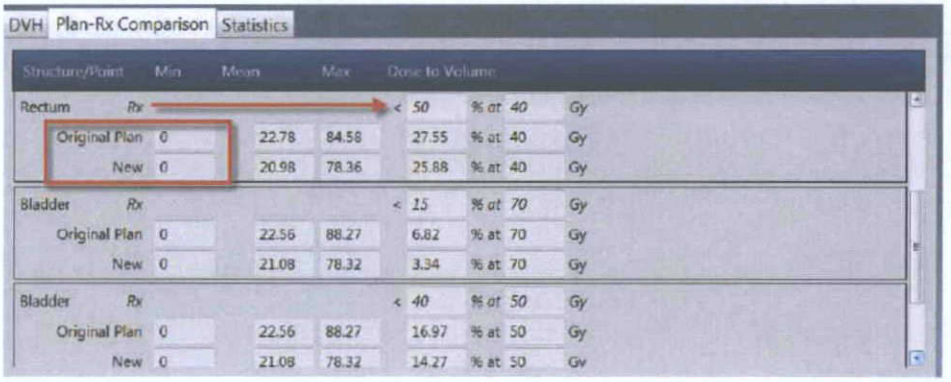
1. **Click > Re-Optimize Dose to New Plan.**
   - ❖ A progress bar will appear.

When the progress bar closes the **New Plan** will be appear as isodose lines on the Daily Image and will be compared with the **Original Plan** in the **DVH, Plan-Rx Comparison,** and **Statistics** displays. See examples below.

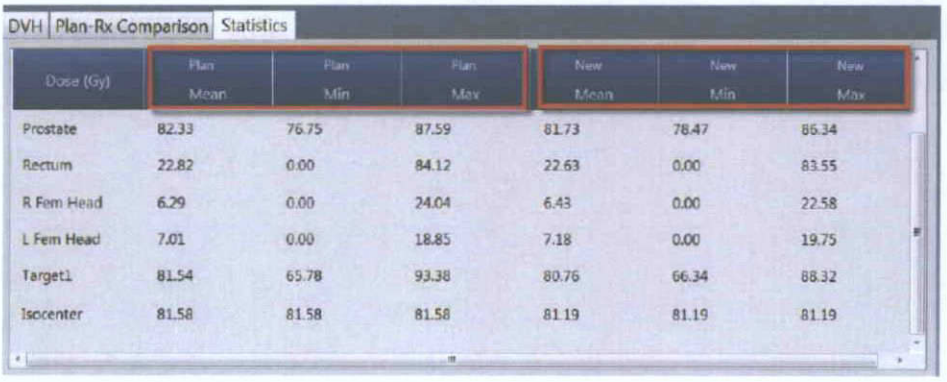
Example of DVH comparison of Plan scheduled for fraction with New Re-Optimized fraction:



Example of Original and New Plan Comparison:



Example of Plan and New Plan Statistics:



Dose (Gy)	Plan Mean	Plan Min	Plan Max	New Mean	New Min	New Max
Prostate	82.33	76.75	87.59	81.73	78.47	86.34
Rectum	22.82	0.00	84.12	22.63	0.00	83.55
R Fem Head	6.29	0.00	24.04	6.43	0.00	22.58
L Fem Head	7.01	0.00	18.85	7.18	0.00	19.75
Target1	81.54	65.78	93.38	80.76	66.34	88.32
Isocenter	81.58	81.58	81.58	81.19	81.19	81.19

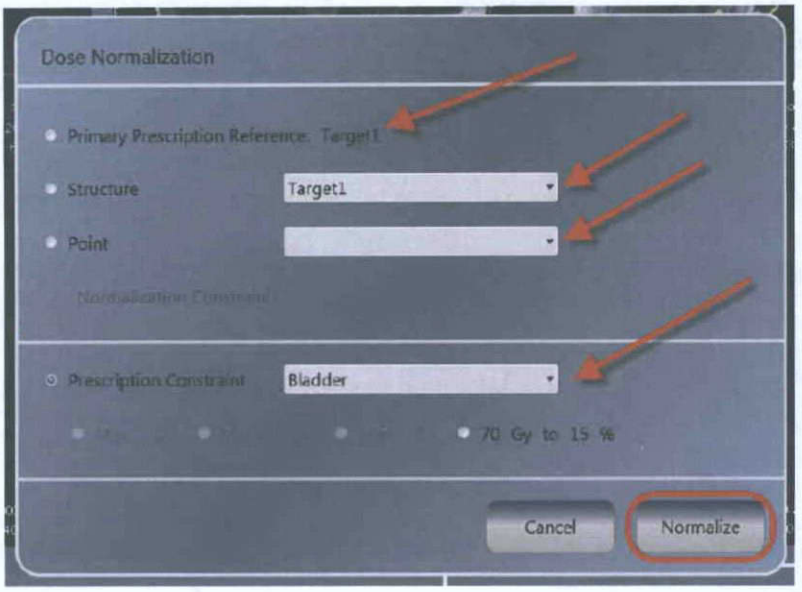
5.8 Normalizing New Plan to Prescription

To meet the prescription treatment planning objectives you may want to normalize the new plan to the prescription. For more information on the principle of dose normalization see "Introduction to Dose Normalization" on page 106.

To begin the normalization process,

1. **Click > Normalize New Plan to Rx** from the **Dose** group box.

❖ An information box opens.



2. **Choose the normalization parameter—Primary Prescription Reference, Structure, Point, or Prescription Constraint.**

3. **Click > Normalize.**

   ❖ The new plan is normalized to the parameter you chose.

5.9 Editing New Plan in Planning Workflow

You can go back to the planning workflow and edit any part of the treatment plan.

1. Click > **Edit New Plan in Planning Workflow** from the **Dose** group box

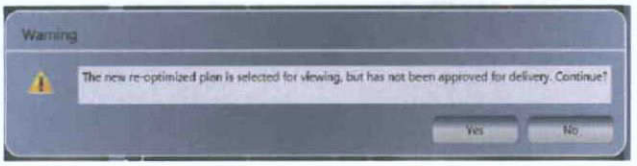
❖ **The Dose screen in the Treatment Planning workflow will open.**

➢ **You can go to any of the enabled screens to edit the plan. There is a Back menu button that will return you to the Treatment Delivery workflow.**

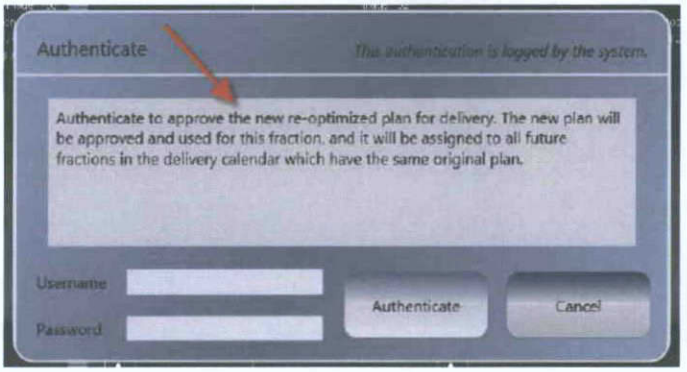
5.10 Treating with Re-Optimized Dose

NOTE: If you **Re-Optimize Dose to New Plan** but do not select **Treat with New Plan** from the **Finalize Treatment Plan** group box, the fraction you deliver will be from the **Original Plan.**

You will receive the following **Warning message** if you have re-optimized and attempt to go to the next part of the delivery workflow. If you choose Continue by clicking the Yes button, you will continue with delivery of the original plan. If you choose No, then you must choose the optimized plan in the Finalize Treatment Plan group box and then authenticate.



Once you authenticate the re-optimized plan, it will replace the original plan on the patient's delivery calendar.



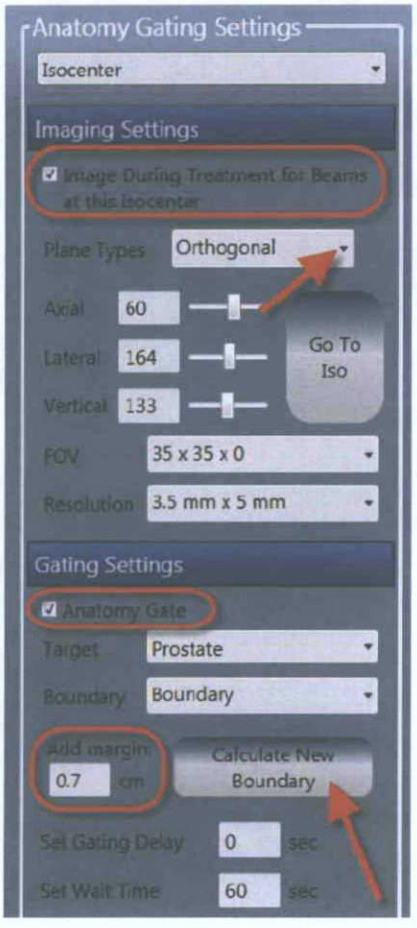
5.11 Anatomy Gating

Anatomy gating is available from both the Treatment Planning and Treatment Delivery interfaces.

It is an optional feature that allows you to set up spatial and time tolerances for pausing a treatment delivery. In other words, you can stop treatment when a structure moves too far outside of a boundary you have set and you can determine how long you want to delay gating once movement is detected and how long you want a paused treatment to "stand by" before ending the treatment early.

The spatial tolerances for anatomy gating are set by selecting a target structure and a boundary structure. Whenever the target structure goes beyond the boundary you set, then treatment is paused until the target moves back into range.

The target structure must be a pre-existing structure, imported or created on the structures screen.



Use the **Imaging Settings** group box to set up the gating image planes.

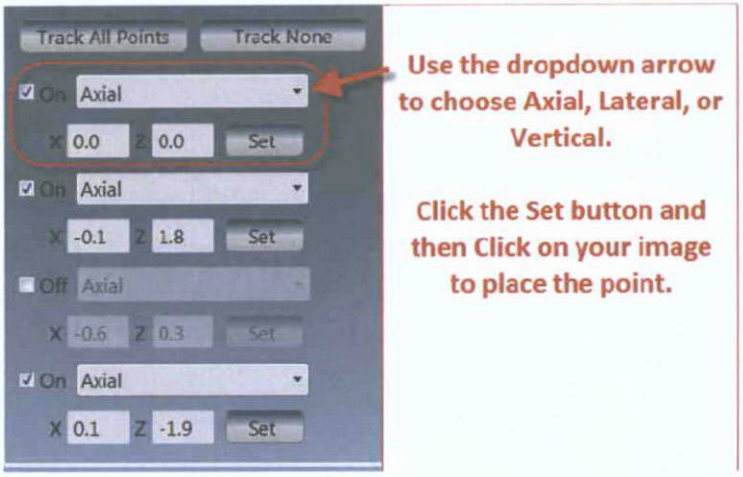
Use the drop down arrow next to **Plane Types** to select either **Axial, Lateral, Vertical** or **Orthogonal.**

Use the sliders to move through the slices to select the images you want to use for gating or use the **Go to Iso** button to move all the slices to Isocenter.

You set the **Target** and **Boundary** structures from the drop down boxes in the **Gating Settings** group box. You can also select a structure and then add a margin (in cm).

5.12 Tracking Points

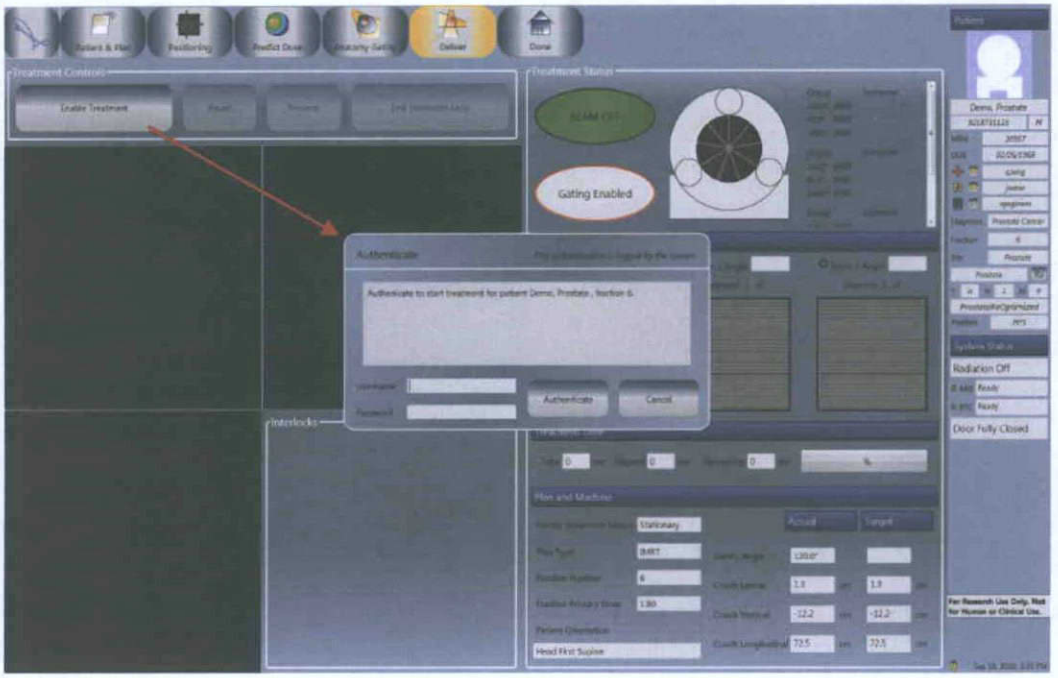
You can set up to four anatomical tracking points and track their 3D motion during delivery. A point can be set at the centroid of a structure or set to track a point on the boundary of two structures. Point tracking records position coordinates in each acquired image and in the Fraction Summary you can view a tracking point chart. See "Viewing the Tracking Points Chart" on page 48.



5.13 Delivering a Treatment Fraction

When you open the **Deliver** interface and select **Enable Treatment** you will be prompted to authenticate the treatment. To actually start the treatment fraction after you **Click > Enable Treatment**, you must press start on the control panel .

Before you authenticate, always check the patient sidebar to be sure that the patient you are treating is the patient identified in the sidebar.



The treatment fraction is delivered in a "step and shoot" method. Each beam segment is delivered separately and the radiation shielded (turned off) while the MLCs move into the shape of the next segment.

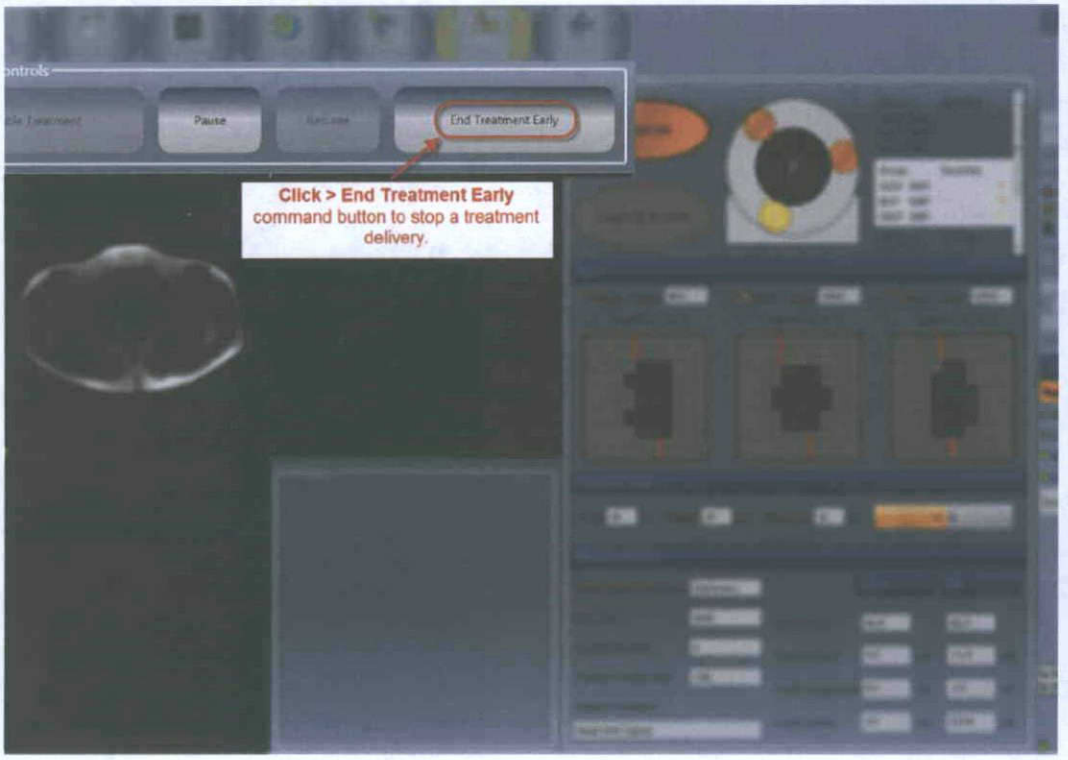
If the treatment fraction you are delivering has gating enabled then the radiation will also pause whenever the target is out of bounds. See "Anatomy Gating" on page 146.

5.13.1 Ending Treatment Early

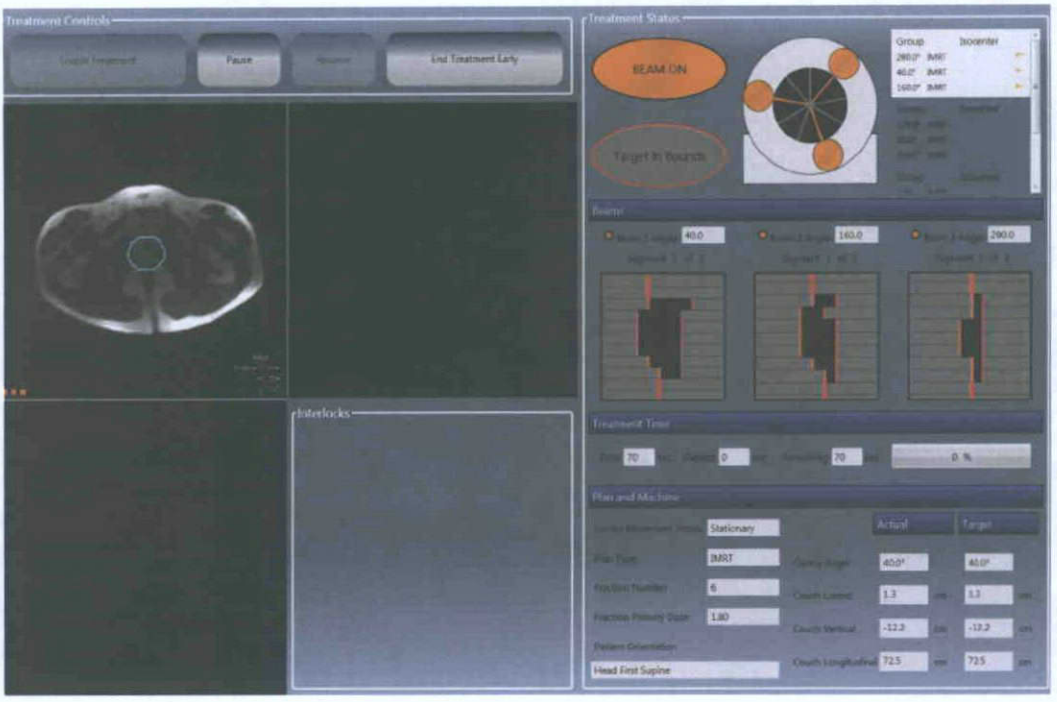
If it is necessary for any reason to stop a treatment before the full fraction has been delivered you can use the **End Treatment Early** command button and the ^{60}Cobalt sources will be retracted and the MLCs closed.

To stop a treatment delivery at any time,

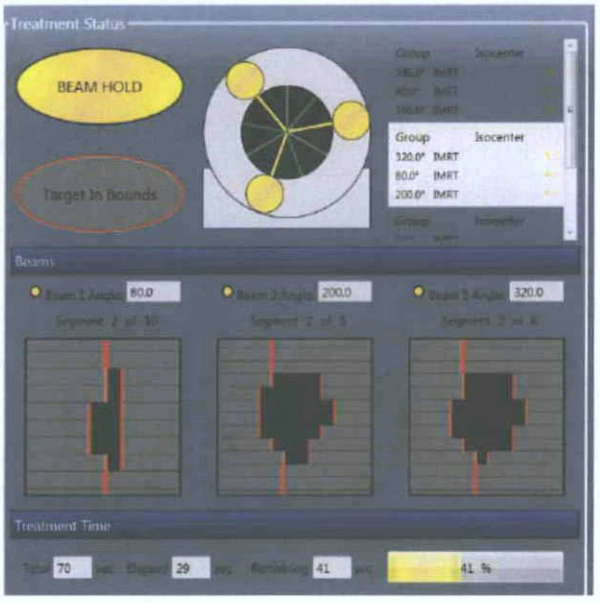
1. **Click > End Treatment Early.**



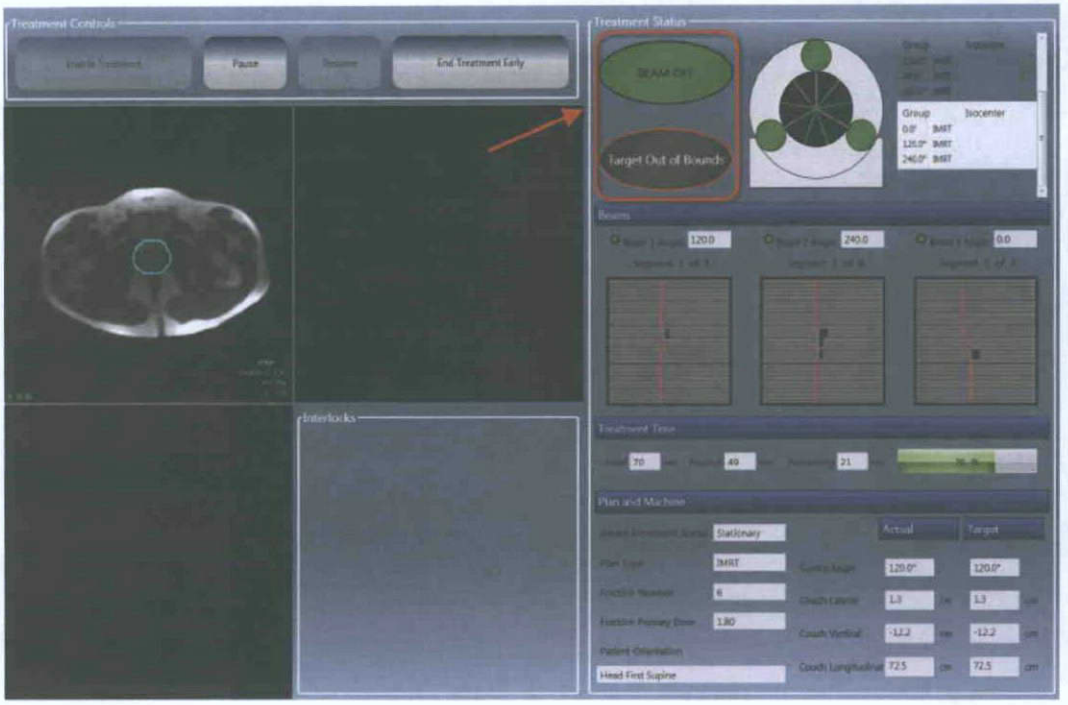
This is the Beam-On state with imaging enabled for one plane (axial).



A Beam Hold state while the MLCs move into place for the next Beam On segment.



A Gating Beam Off state when the target is out of bounds.



A Beam On state when imaging has been disabled.

THIS PAGE INTENTIONALLY LEFT BLANK

Chapter 6: Quality Assurance

6.1 Overview of QA Tests and Procedures

This chapter contains step-by-step instructions for the ViewRay System™ quality assurance procedures. The American Association of Physicists in Medicine considers it imperative that an appropriate quality assurance program be implemented. To help you fulfill that AAPM recommendation, the ViewRay System™ is equipped with the following QA screens:

- Plan QA
- SNR/Uniformity
- Spatial Integrity
- Image Localization
- Output Factors
- Source Tracking
- Gating Latency
- MLC Test
- Delivery Records

6.1.1 Recommended Frequencies for QA Tests

TABLE 6-1 Daily ViewRay System™ QA Tests

TEST	TYPE/TOLERANCE
Verification of Door Interlock	Functional
Radiation Room Monitor	Functional
Audio-Visual monitor	Functional
Lasers	≤ 2mm
Imaging and treatment coordinate coincidence	≤ 2mm
Positioning/Repositioning	≤ 1mm

TABLE 6-2 Weekly ViewRay System™ QA Tests

TEST	TYPE/TOLERANCE
Check MLC Shape and Compare to Expected Output	Visual inspection for discernable deviations such as an increase in interleaf transmission
Check Source Positioning and Timing	3 mm

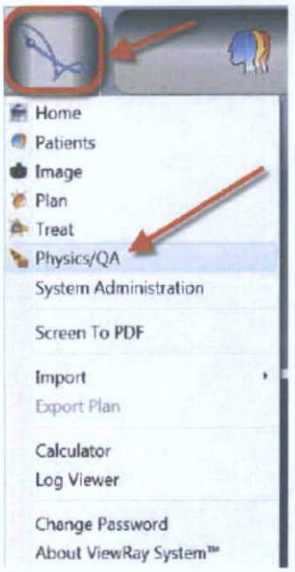
TABLE 6-3 Monthly ViewRay System™ QA Tests

TEST	TYPE/TOLERANCE
MLC Settings v. Radiation Field for Two Patterns (non-IMRT)	2 mm
MLC Travel Speed	Loss of leaf speed > 0.5 cm/s
Leaf Position Accuracy	1 mm for IMRT leaf positions for 2 gantry angles on each head (Picket Fence may be used)
Dosimetry Output	2%
Radiation Field MRI Coincidence	3mm
Field Size Indicator	2mm
Gantry Angle Indicator	1 degree
Emergency Off	Functional
MRI Geometric Distortion	Spatial Integrity Phantom
MRI Spatial Resolution	Spatial Integrity Phantom
MRI Contrasts	Siemens Phantom
MRI Uniformity and Signal-to-Noise Ratio	Siemens Phantom
Gating Latency	ViewRay Gating Phantom

TABLE 6-4 Annual ViewRay QA Tests

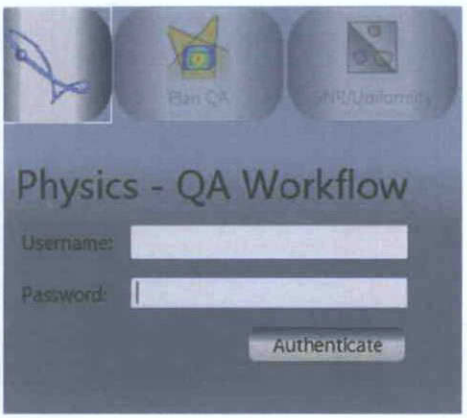
TEST	TYPE/TOLERANCE
Dosimetry Output	2%
Field Size Dependence of Output Consistency	2%
Central Axis PDD Consistency	2%
Transmission of Table, Patient Coils, and Accessories	2%
Timer Linearity and Error	1%
Output Constancy with Gantry Angle	2%
Safety Interlocks	Functional
Gantry Rotation Isocenter	2mm diameter
Coincidence of Gantry and Couch	2mm diameter
Couch Top Sag	2mm
Vertical Table Travel	2mm
MLC Transmission (Average of Leaf and Interleaf Transmission)	± 0.5% of baseline
Leaf Position Repeatability	± 1.0mm
MLC Spoke Shot	≤ 1.0mm radius
Coincidence of MRI and Radiation Field	± 2.0mm
Segmental IMRT Step-and-Shoot Test	< 0.35cm max. error RMS, 95% of error counts <0.35 cm

6.2 Accessing the QA Screen



To access the QA Screen:

1. **Click > ViewRay Icon** on the menu bar.
   - ❖ A drop down menu will appear.
2. **Click > Physics/QA**
   - ❖ The **Physics - QA Workflow** Screen will open and you will be prompted to enter your username and password.



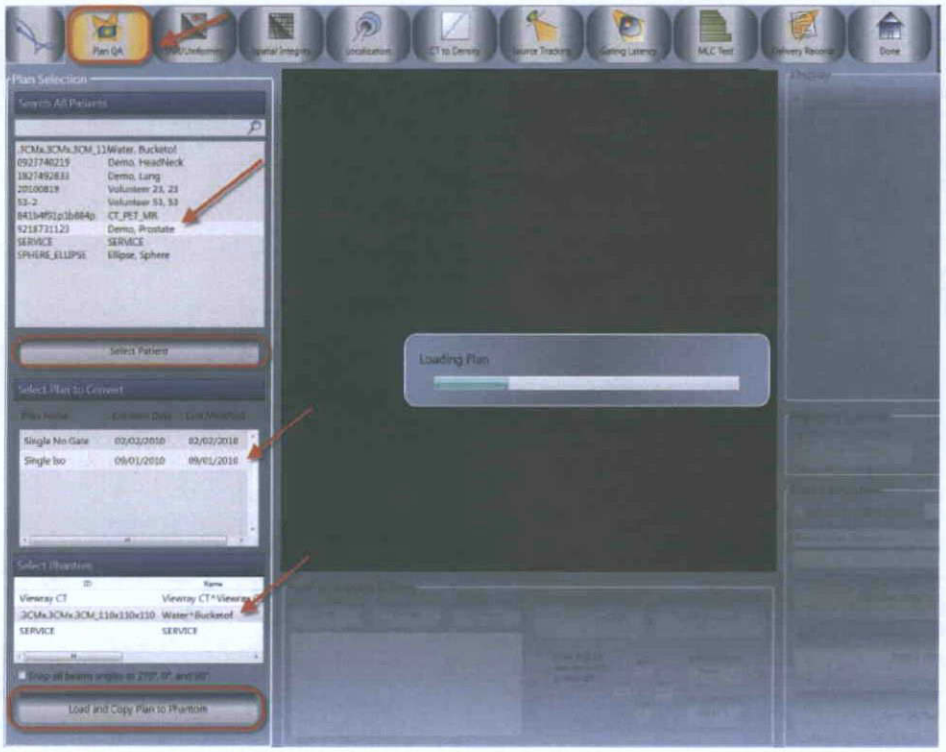
Enter your Username and Password and Click > Authenticate.

6.3 Plan QA

IMRT plans approved for delivery to a patient should be tested on a phantom to measure the accuracy of the isodose distribution as compared to the prescribed dose.

To begin a Plan QA:

1. **Click > Plan QA icon** from menu bar.
   - ❖ The Plan QA interface opens.
2. **Click on a patient** from the Patient List
3. **Click > Select Patient.**
   - ❖ All approved plans for that patient will appear in the **Select Plan to Convert** list.
4. **Click** on a plan from the list.
5. **Click** a phantom from the **Select Phantom** list.
6. **Click > Load and Copy Plan to Phantom.**
   - ❖ The **Loading Plan** progress bar appears.

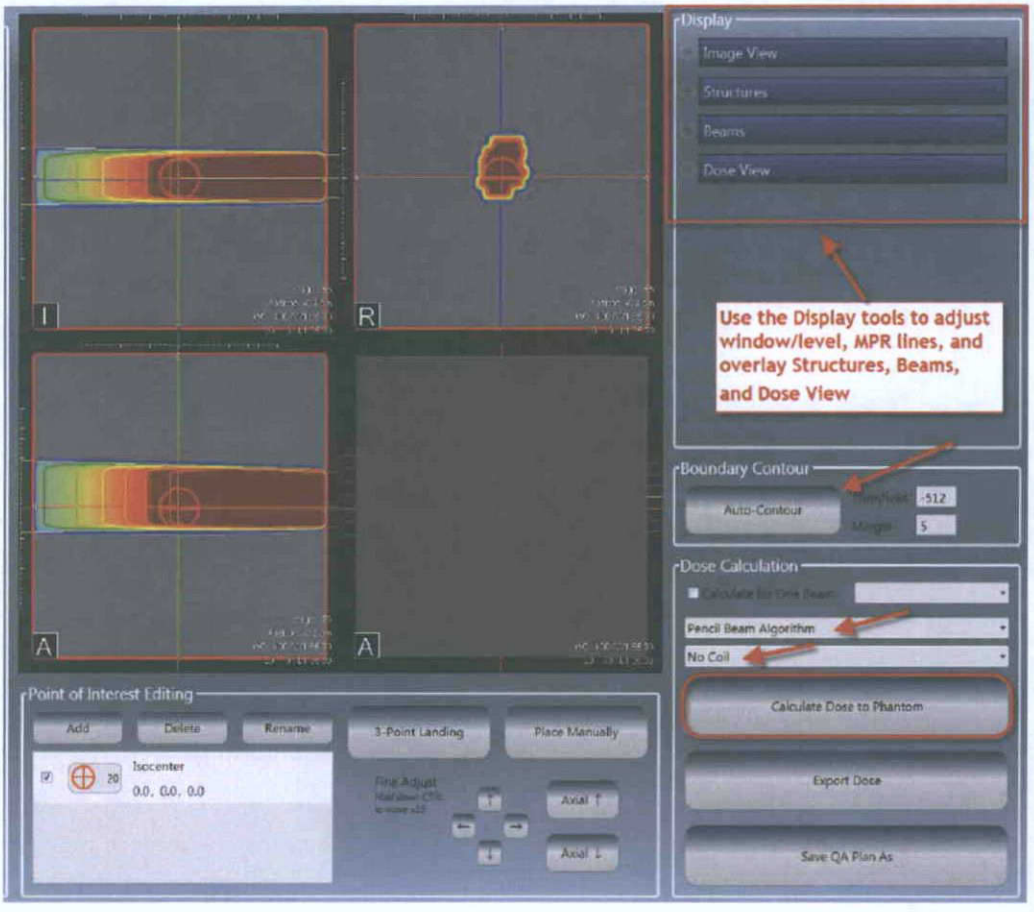


6.3.1 Calculating Dose to Phantom

After loading an approved plan to a phantom, you can change the phantom boundary, adjust viewing options, edit points of interest, choose **Pencil Beam** or **Monte Carlo** algorithms for the dose calculation, and choose **Coil** or **No Coil.**

When you are satisfied with your choices,

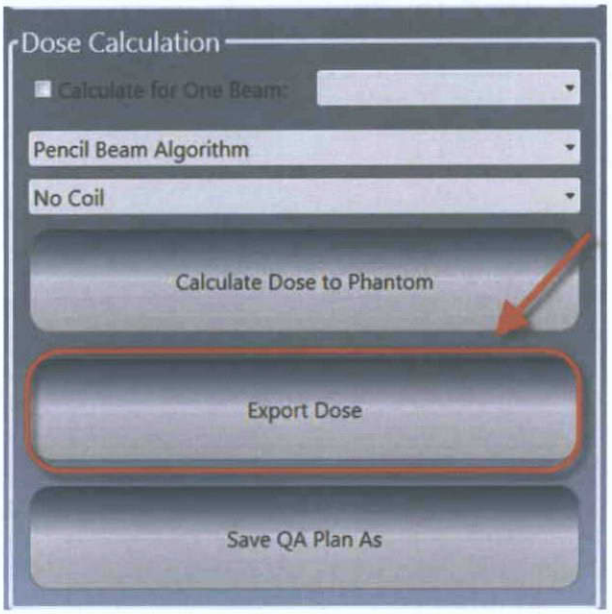
Click > Calculate Dose to Phantom.



6.3.2 Exporting Dose

To export the dose,

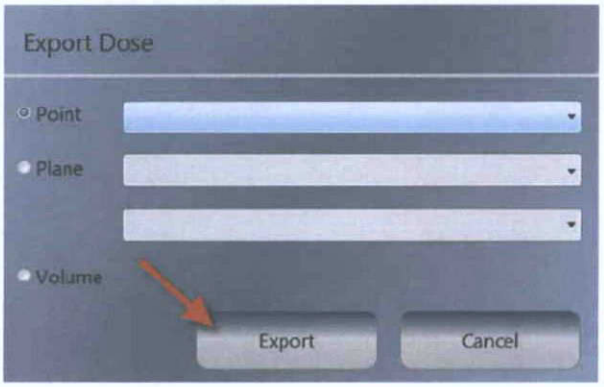
1. **Click > Export Dose** from the Dose Calculation group box.



❖ The **Export Dose** information box will open (pictured below).

2. Select **Point, Plane,** or **Volume.**

3. **Click > Export.**

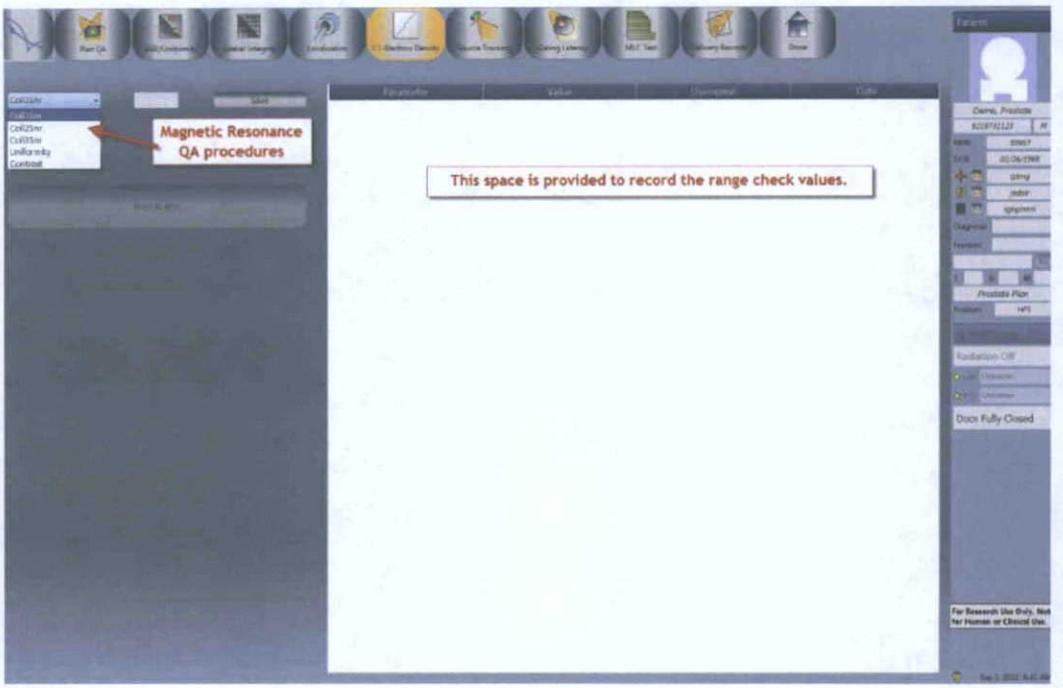


❖ The **Save As** information box will appear.

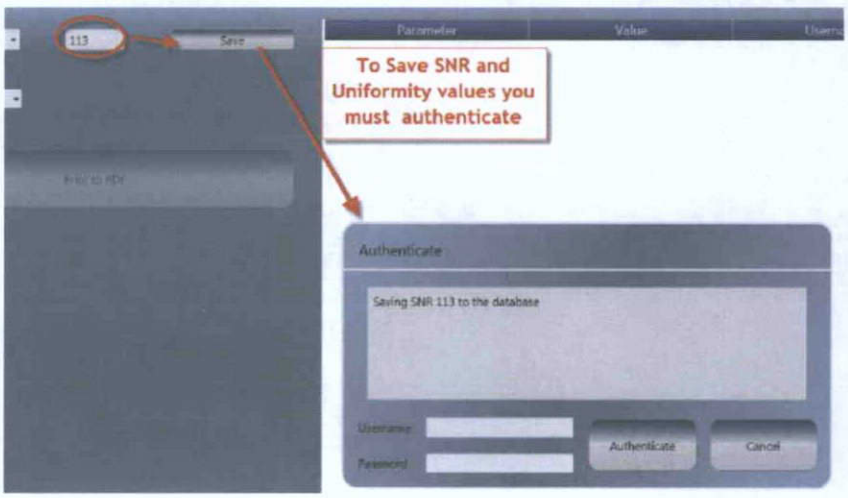
4. Indicate which folder and under what name you want the Exported Dose saved.

6.4 SNR/Uniformity

This QA interface provides a place to record (and save to PDF) the data from the signal-to-noise, uniformity and contrast magnetic resonance QA procedures (described in the MR manual). The data recorded here is a range check to ensure that the data collected from the QA tests are within normal operating ranges.

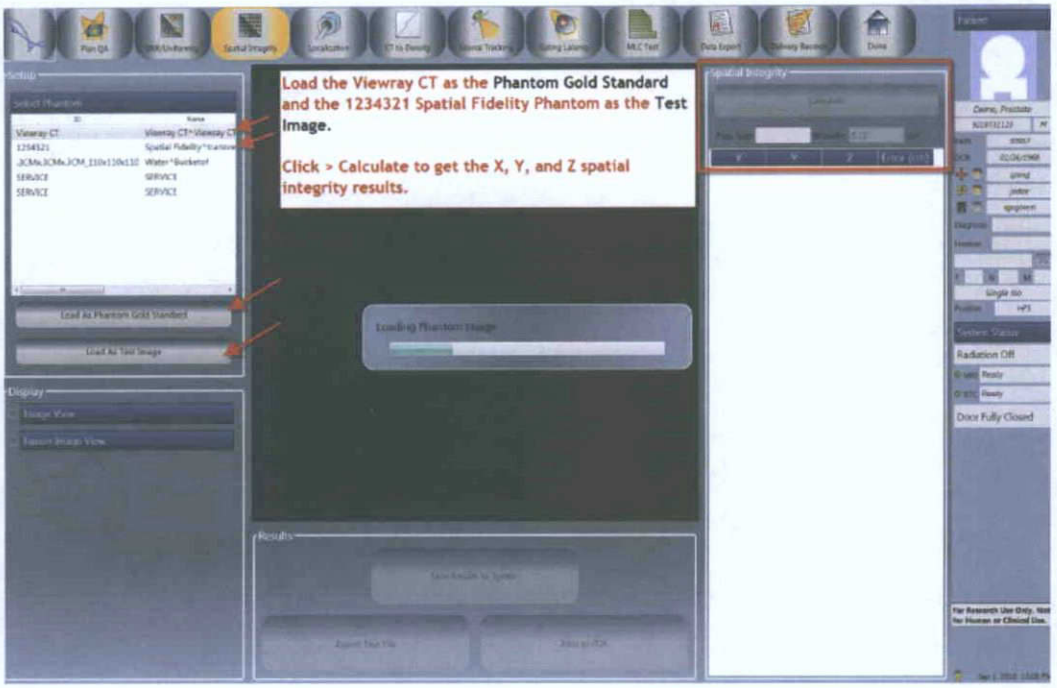


Enter a value and then select **Save**, you will be prompted to authenticate.



6.5 Spatial Integrity

Inhomogeneities in the magnetic field can result in geometrically distorted images. This QA application is for verifying the spatial integrity of the system's MR images.



To run a spatial integrity test:

1. **Click > Viewray CT** from the **Select Phantom** list.
2. **Click > Load as Phantom Gold Standard** command button.
   - ❖ The **Loading Phantom Image** progress bar will appear. When it disappears begin step 3.
3. **Click > 1234321 Spatial Fidelity** from the **Select Phantom** list.
4. **Click > Load as Test Image.**
   - ❖ The **Loading Phantom Image** progress bar will appear. When it disappears begin step 5.
5. **Click > Calculate** command button from the **Spatial Integrity** group box.
   - ❖ The X, Y, and Z centroid offsets appear in the column on the left. See screenshot on next page.

Spatial Integrity Results Screen

All spatial integrity offset errors 0.15 and greater are highlighted in yellow.

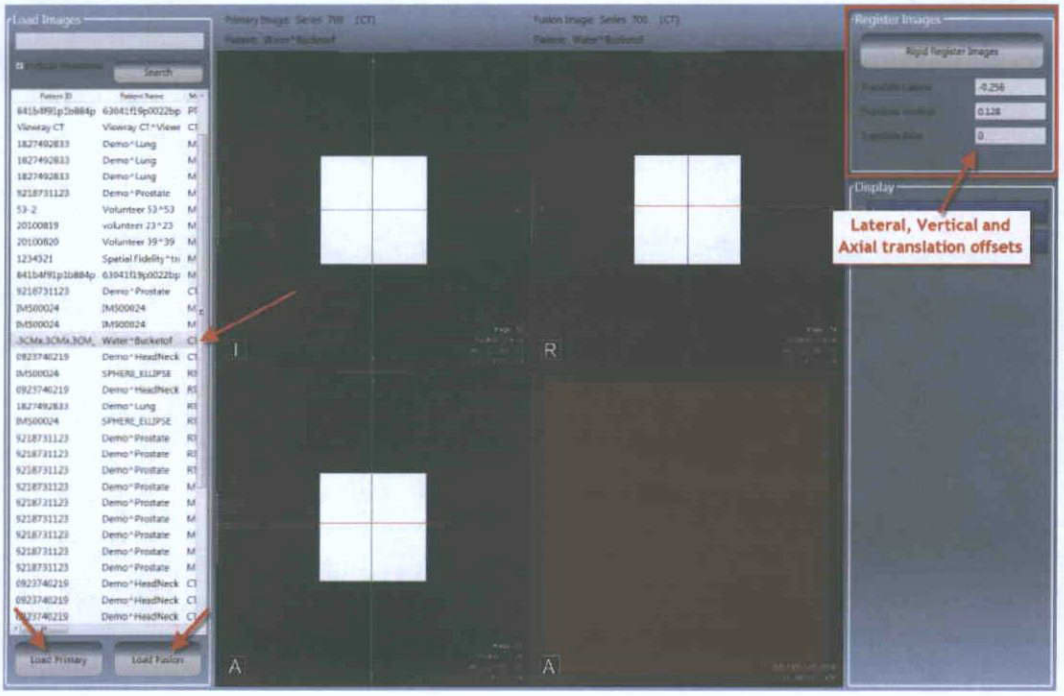
Call ViewRay Service for all offset errors that are highlighted in yellow.

6.6 Localization

This QA interface is designed to measure couch repositioning repeatability and couch sag. AAPM guidelines recommend that the couch should be checked annually for localization repeatability and sag and that both should be within a tolerance of 2mm.

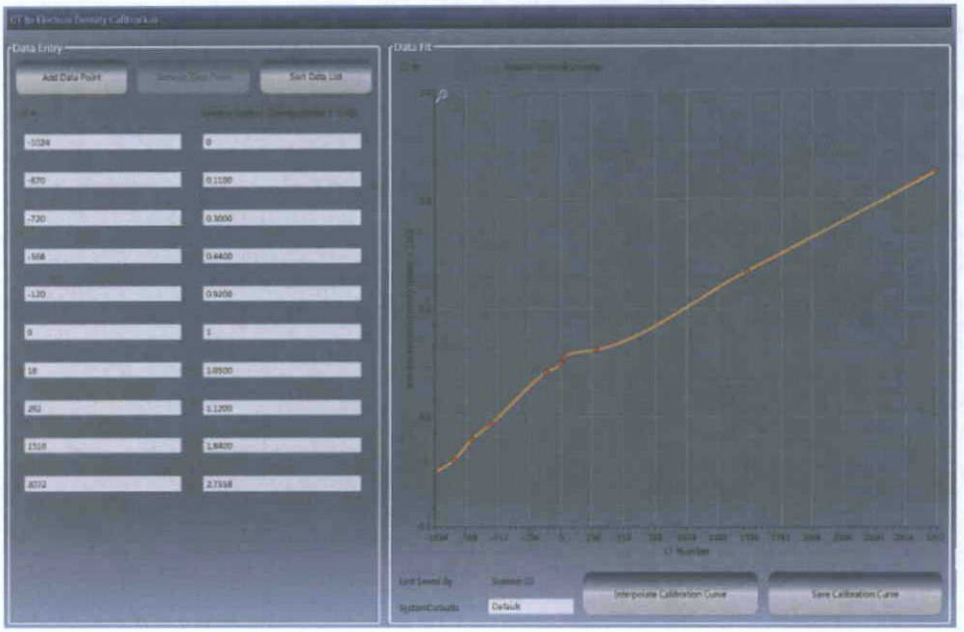
To perform a couch sag test:

1. Select the **Water^Bucketof** Phantom from the **Load Images** list.
2. **Click > Load Primary** to load the phantom as the primary image.
3. Select the **Water^Bucketof** Phantom from the **Load Images** list.
4. **Click > Load Fusion** to load the phantom as the fusion image.
5. **Click > Rigid Register Images.**
   - ❖ The **Lateral, Vertical** and **Axial** translation offsets will appear in the boxes beneath **Rigid Register Images.**



6.7 CT to Electron Density

To precisely account for tissue heterogeneities and for the system to accurately predict dose, a radiotherapy treatment planning system must have a conversion table for Hounsfield units (HU) to electron density.



Houndsfield units are used to distinguish the amount of x-ray attenuation of each voxel in a 3D image. The voxels are represented as 12-bit binary numbers and therefore have 12^{12} = 4096 possible values. These values are arranged on a scale from -1024 HU to +3072 HU.

The range in HU is from air to dense bone and is entered and plotted on the look-up table. Once the look-up table is configured, the system references the table and automatically generates the electron density map for the selected CT images.

WARNING: The table must have an endpoint (highest value). If it does not, then any values that are higher than your table's endpoint will be assigned to the electron density setting of the last value on your table. This could result in a miscalculation of dose.

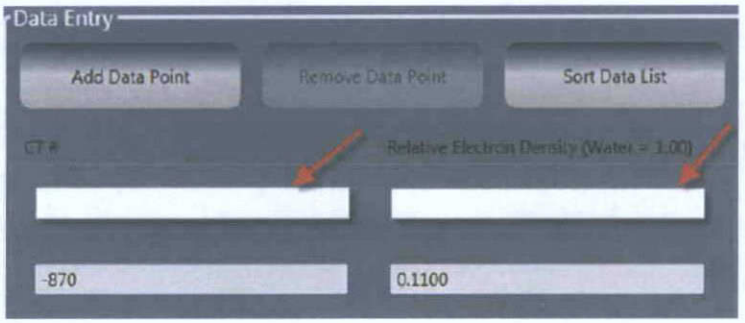
6.7.1 Adding Data Points to Look-Up Table

Because data point additions to the CT to Electron Density look-up table affect dose calculations for plans using CT, all additions to the look-up table must be authenticated.

To add a data point,

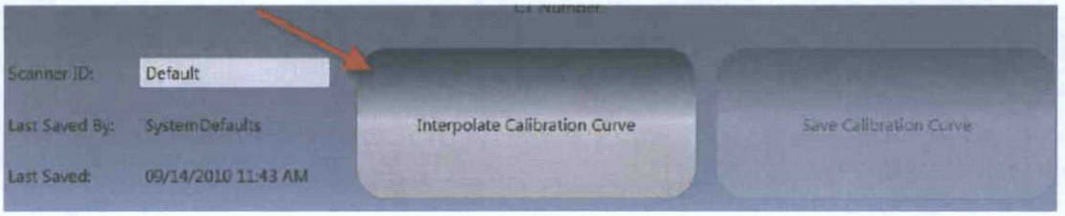
1. **Click > Add Data Point.**

❖ Two blank lines will appear in the **Data Entry** group box.

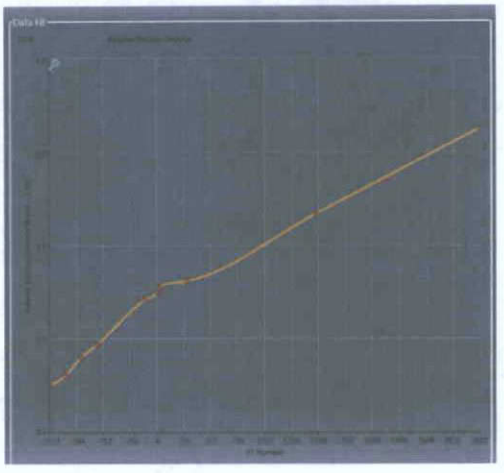


2. Add your CT to Electron Density plot points.

3. **Click > Interpolate Calibration Curve**

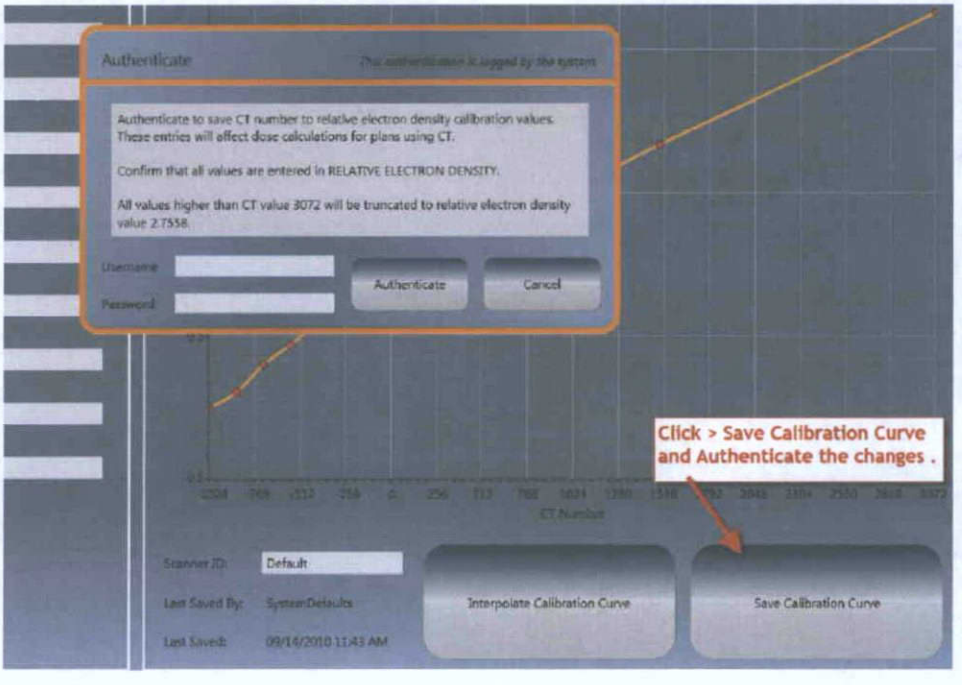


❖ The Calibration Curve is plotted.



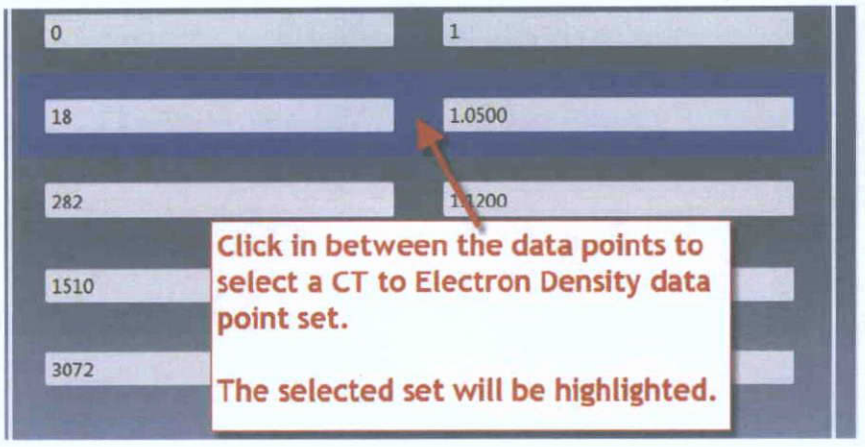
To save the new Calibration settings,

1. Click > Save Calibration Curve.
   ❖ An Authentication box opens.
2. Authenticate to save the new CT to Electron Density calibration settings.



6.7.2 Deleting Data Points from the Look-up Table

1. Click in between a set of data points.



2. **Click > Remove Data Point.**

3. **Authenticate to save your changes.**

6.8 Source Tracking

The three ^{60}Cobalt sources are calibrated when they are loaded. The system keeps track of the source strength and uses it in all dose calculations and optimizations.

ONLY CERTIFIED PERSONNEL ARE AUTHORIZED TO MAKE CHANGES ON THE SOURCE TRACKING INTERFACE.

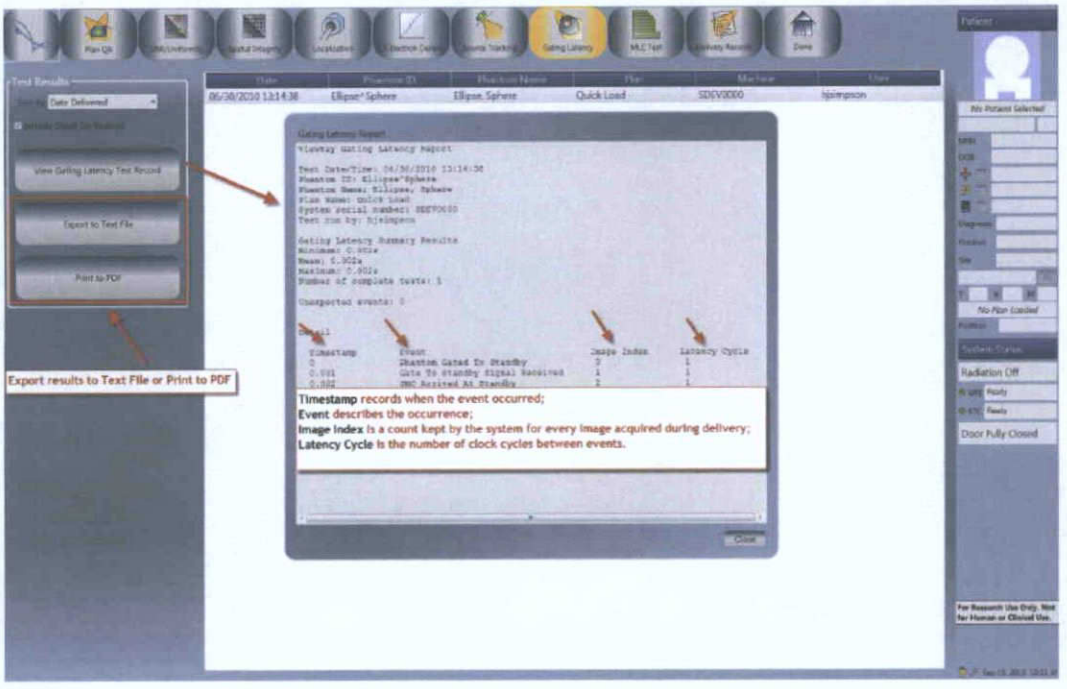
6.9 Viewing the Gating Latency Test Report

The system captures and records gating latency test results.

To view the test results,

1. Click > View Gating Latency Test Record.

You can also export the results to a text file or print to PDF.



6.10 Running an MLC Speed Test

The MLC speed test is designed to test the condition and speed of the MLCs. The MLCs are moved to designated positions in a specified manner to determine if the MLCs are in working order.

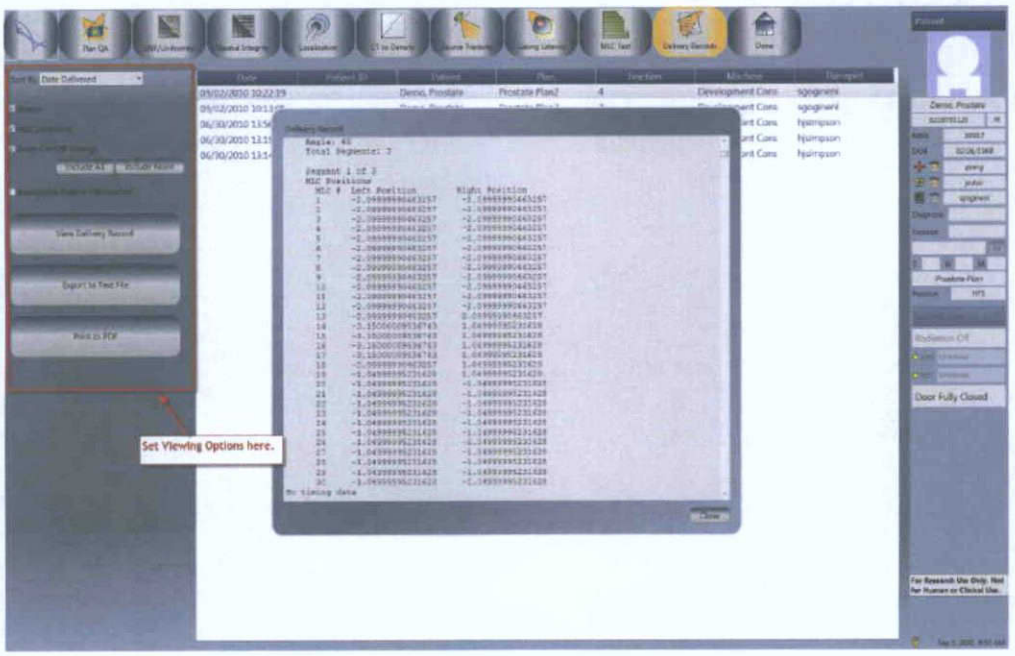
If the speed is detected to be outside a specified range, that particular speed data is highlighted in yellow. When any speed data is highlighted in yellow the operator should call service before using the system for treatment delivery.

The following screenshot shows an example of MLCs with speed data out of specification.

6.11 Viewing Delivery Reports

This QA interface provides Delivery Reports that can include beams, MLC locations, beam on-off timings. There is a setting to anonymize the patient information.

You can export the reports to text files and view and print to PDF.



ViewRay System™ ERROR MESSAGES

ID	DIALOG MESSAGE	USER ACTION
SY00001	The system found plan recovery data. Would you like to try to recover the last plan?	Decide whether the recovered plan is to be restored or not.
SY00002	Message key was not found in the resource file.	Dismiss dialog. Report message to service.
DA00001	Pre-save check failed.	Re-load the plan. The work done since the last save is lost.
DA00002	{0} [Specific Message will come from PACS]	Re-try the export, or examine the PACS logs to determine the cause.
DA00003	Could not connect due to unknown destination.	Check setup of PACS.
DA00004	Structure ROI #{0} does not have a name.	Name the structure before exporting the structure. The re-import the structures.
DA00005	ROI Contour contains reference to unknown structure number #{0}.	Try re-exporting the structures from the source system and then re-import the structures.
DA00006	Contour number {0} in sequence number {1} for structure {2} is not closed co-planar. Only close co-planar contours are supported. These contours will be skipped.	Re-export the contour as co-planar or manually draw the contour.
DA00007	Contour number {0} in sequence number {1} for structure {2} is spans multiple slices. Only contours on axial slices are supported. This contour will be skipped.	Export the contours using axial images or redraw the contours manually.
DA00008	The structures in this file are not defined in the space of the primary image set.	Check to make sure the correct contours are exported, the re-import the contours.
DA00010	Not all contours lie directly on a slice. Contours will be shifted by up to {0}cm onto the nearest slice.	Check to ensure the contours are imported as desired.
DA00011	The specified file does not exist.	Check the filename and reselect the file.
DA00012	The specified file is not a valid DICOM format.	Check to ensure the file is exported in DICOM format and then try to re-import the file.
DA00013	The specified file does not conform to the DICOM standard. See the log for details.	Modify the file to correct the issue, and try again.
DA00014	A plan with that name already exists.	Choose a diferent filename for saving the plan.
DA00015	A contour template with that name already exists.	Attempt to save the template under a different name.
DA00016	A patient with id '{0}' already exists.	Choose a different patient ID before attempting to save the new patient information.
DA00017	The modality of the data is {0} which is not supported for this operation.	Only import CT, PETE, and MRI images.

ID	DIALOG MESSAGE	USER ACTION
DA00018	The file {0} already exists.	Attempt to save the file under a different name.
DA00019	Could not connect to the database. Ensure that the database is running and that the wires are connected properly. If the problem persists, contact service.	Check network connections and database server. Call Service.
DA00020	Could not connect to the import service. Ensure that the service is running, and that the connection settings are correct. If the problem persists, contact service.	Reboot the system. If the problem persists, call Service.
DA00021	The series does not meet the requirements of the system.	Choose a series that has DICOM reqired fields for import.
DA00022	The DICOM AE title already has an entry in the registry.	Change the AE title and attempt to save again.
DA00023	The source tracking data is corrupted. Contact service.	Contact service.
DA00024	The head number field is corrupted. Contact service.	Call service.
DA00025	There was an error retrieving source records. Contact service.	Call service.
DA00026	There was an unknown error checking the file.	Retry the operation, if the problem persists, call service.
DA00027	The system could not query the PACS. Check your connection information, and ensure that the PACS is available.	Check the network connections and the PAC status.
DA00028	Could not assign plan to fraction. One or more selected fractions has been delivered.	Choose an unallocated fraction for assignment.
DA00029	Cannot unassign plan to fraction number {0}. The fraction has already been delivered.	Choose another fraction to unassign.
DA00030	Cannot lock fraction number {0}. The fraction has already been delivered.	Select another fraction to lock.
DA00031	Unable to remove the plan '{0}' because it has been used for treatment, and has not been archived.	Archive the plan and then attempt to delete.
DA00032	Unable to remove the plan '{0}' because it is assigned on the Delivery Calendar. You must unassign the plan from the fraction before deleting, or you may delete the entire diagnosis.	Un-assign the plan, and try again.
DA00033	Unable to remove plan {0} because it is readonly.	Do not attempt to delete readonly plans.
DA00034	Unable to remove plan {0} because it is currently open.	Retry after the plan is closed.

ID	DIALOG MESSAGE	USER ACTION
DA00035	Unable to un-approve plan. The plan has been delivered.	Go to delivery calendar and unassign the remaining fractions for the plan.
DA00036	A user with username '{0}' already exists.	Choose another username and try again.
DA00037	None or multiple users with username '{0}' exist, Count = '{1}'.	Call service, the database has been compromised.
DA00038	Password change failed. Check your username and password and retry. If this error persists, contact service.	Re-enter username and password and retry.
DA00039	The timing data for this treatment delivery could not be loaded. Try again without beam on-off timings.	To see available data, turn off beam on-off timing selection box and call service to report the problem.
DA00040	A template with that name already exists. Please enter a different template name.	Try a different name and save again.
DA00041	Dose series must have Dose Type of Physical or Effective, and must be in Gy.	Only attempt to import dose file with required DICOM fields.
DA00042	Unable to retrieve users from database.	Call service.
DA00043	Unable to retrieve beam template data from database.	Call service.
DA00044	Unable to save beam template data to database.	Retry saving the template. If the problem persists, call service.
DA00045	Unable to delete beam template from database.	Try to delete the template. If the problem persists, call service.
DA00046	You cannot save this plan. This may be because the plan is open by another user, because it is approved, or because it has been used for delivery. To save changes, please save the plan as a new plan.	Use "Save As" to save the plan under a different name.
DA00047	The file '{0}' is not a valid plan archive file, or the file has been corrupted.	Choose the correct file or choose another archive to restore.
DA00048	The patient id in the file is already in use by a patient that does not match on name, DOB, or sex.	Do not reuse patient IDs on multiple systems.
DA00049	This series cannot be deleted because the following plans reference it. {0}	Remove the images from the plan or archive/delete the plan.
DA00050	There are {0} images in series {1} that are not referenced by this plan. These images have not been loaded. This may indicate that images were added to the series after this plan was created.	Reimport the planning images. Confirm the entire series imported into the system before adding to a plan.
DA00051	The image set could not be loaded because it contains an invalid value.	Import a scan with valid values.
DA00052	The image set could not be loaded because it contains a value that is not acceptable for radiotherapy planning.	Import a scan that is acceptable for radiotherapy planning.

ID	DIALOG MESSAGE	USER ACTION
DA00053	The plan contains invalid data. This may indicate plan corruption.	Re-load the plan, or create a new plan.
DA00054	The plan contains invalid data. This may indicate plan corruption. Check that all of the plan data is valid, and try again.	Check that all of the plan data is reasonable, and try saveing again.
DA00055	The system could not connect to the Remote Import Service. Ensure that the service is running and try again. If the problem persists, call service.	Check that the system/service is running. Service should check that the configuation is correct if the problem persists.
DA00056	You can only combine series that are for the same patient.	Select two or more series that are from the same patient and try again.
DA00057	You can only combine series that are in the same modality.	Select two or more series that are in the same modality and try again.
DA00058	You can only combine series that are MRI or CT.	Select two or more series that are in a valid modality and try again.
SV00001	WCF Services Exception.	An error in the delivery system has occurred. Call Service.
SV00002	Unknown Message.	An error in the delivery system has occurred. Call Service.
SV00003	Unhandled Exception in Services.	An error in the delivery system has occurred. Call Service.
SV00004	Could not open RTC Session.	An error in the delivery system has occurred. Call Service.
SV00005	Could not close RTC Session.	An error in the delivery system has occurred. Call Service.
SV00006	No communication with services.	An error in the delivery system has occurred. Call Service.
SV00007	Could not send the heartbeat to RTC.	An error in the delivery system has occurred. Call Service.
SV00008	Could not execute heartbeat services on console.	An error in the delivery system has occurred. Call Service.
SV00009	Could not allocate memory for interlocks.	An error in the delivery system has occurred. Call Service.
SV00010	Could not read interlocks.	An error in the delivery system has occurred. Call Service.
SV00011	Could not execute services interlocks on console.	An error in the delivery system has occurred. Call Service.
SV00012	Could not open treatment delivery service session.	An error in the delivery system has occurred. Call Service.
SV00013	Could not close treatment delivery service session.	An error in the delivery system has occurred. Call Service.
SV00014	Could not start treatment delivery.	An error in the delivery system has occurred. Call Service.

ID	DIALOG MESSAGE	USER ACTION
SV00015	Could not stop treatment delivery.	An error in the delivery system has occurred. Call Service.
SV00016	Could not get gating plane visualization data.	An error in the delivery system has occurred. Call Service.
SV00017	Could not subscribe for visualization data.	An error in the delivery system has occurred. Call Service.
SV00018	Exception occurred during gating signal change. Sending 'End Early' command to RTC.	An error in the delivery system has occurred. Call Service.
SV00019	Exception occurred during getting treatment delivery result from RTC.	An error in the delivery system has occurred. Call Service.
SV00020	Exception occurred querying treatment delivery result data from RTC.	An error in the delivery system has occurred. Call Service.
SV00021	Console has failed re-connecting with Service Manager. Console and services boxes must be restarted. If problem persists contact service.	An error in the delivery system has occurred. Reboot the system. If the problem persists, call service.
SV00022	Console failed instantiating the right side bar controller.	An error in the delivery system has occurred. Call Service.
SV00023	Console is disconnected from Treatment Delivery Service. Service was closed.	An error in the delivery system has occurred. Call Service.
SV00024	Console has lost communication with Treatment Delivery Service due to Treatment delivery service abort.	An error in the delivery system has occurred. Call Service.
SV00025	Console has lost communication with RTC Couch Service due to unexpected failure.	An error in the delivery system has occurred. Call Service.
SV00026	Console cannot communicate with Treatment Delivery Service because Treatment Delivery Service client is invalid. Call Service.	An error in the delivery system has occurred. Call Service.
SV00027	RTC System has faulted.	An error in the delivery system has occurred. Call Service.
SV00028	RTC System is disconnected from RTC Service.	An error in the delivery system has occurred. Call Service.
SV00029	MRI Scanner System is disconnected from MRI Scan Service.	An error in the imaging system has occurred. Reboot the system. If the problem persists, call Service.
SV00030	MRI Scan Service is disconnected from Data Processor.	An error in the imaging system has occurred. Reboot the system. If the problem persists, call Service.
SV00031	Data Processor service has faulted during data processing.	An error in the delivery system has occurred. Call Service.
SV00032	Data Processor is disconnected from MRI Scan Service.	An error in the delivery system has occurred. Call Service.

ID	DIALOG MESSAGE	USER ACTION
SV00033	Treatment Delivery Service could not be started. Failed connecting to Service Manager.	An error in the delivery system has occurred. Call Service.
SV00034	Treatment Delivery Service failed opening delivery session. Invalid protocol data received.	An error in the delivery system has occurred. Call Service.
SV00035	Treatment Delivery Service cannot process request to due to {0}.	An error in the delivery system has occurred. Call Service.
SV00036	Failed connecting to RTC Service, try restarting the RTC Service.	An error in the delivery system has occurred. Call Service.
SV00037	Failed connecting to MRI Service, try restarting the MRI Service.	An error in the imaging system has occurred. Call Service.
SV00038	Failed connecting to Volume Processor Service, try restarting the Data Processor Service.	An error in the delivery system has occurred. Call Service.
SV00039	Treatment Delivery Service failed opening delivery session for treatment record Id = {0}.	An error in the delivery system has occurred. Call Service.
SV00040	Treatment Delivery Service failed starting delivery for treatment record Id = {0}.	An error in the delivery system has occurred. Call Service.
SV00041	Treatment Delivery Service failed pausing delivery for treatment record Id = {0}.	An error in the delivery system has occurred. Call Service.
SV00042	Treatment Delivery Service failed resuming delivery for treatment record Id = {0}.	An error in the delivery system has occurred. Call Service.
SV00043	Treatment Delivery Service failed stopping delivery for treatment record Id = {0}.	An error in the delivery system has occurred. Call Service.
SV00044	Treatment Delivery Service failed closing delivery for treatment record Id = {0}.	An error in the delivery system has occurred. Call Service.
SV00045	Treatment Delivery Service failed getting delivery data for treatment record Id = {0}.	An error in the delivery system has occurred. Call Service.
SV00046	Treatment Delivery Service failed deleting delivery data for treatment record Id = {0}.	An error in the delivery system has occurred. Call Service.
SV00047	Treatment Delivery Service failed getting imaging data for treatment record Id = {0}.	An error in the delivery system has occurred. Call Service.
SV00048	Treatment Delivery ended early due to an RTC system fault.	An error in the delivery system has occurred. Call Service.
SV00049	Treatment Delivery ended early due to data processing service fault.	An error in the delivery system has occurred. Call Service.
SV00050	Treatment Delivery ended early. Data Processor Service has lost communication with MRI Scan Service.	An error in the delivery system has occurred. Call Service.

ID	DIALOG MESSAGE	USER ACTION
SV00051	Treatment Delivery ended early. MRI Scan Service has lost communication with Data Processor Service.	An error in the delivery system has occurred. Call Service.
SV00052	Treatment Delivery ended early. MRI Scan Service has lost communication with MRI Scanner System.	An error in the imaging system has occurred. Reboot the system. If the problem persists, call Service.
SV00053	Treatment Delivery Service faulted due to unexpected RTC system disconnection.	An error in the delivery system has occurred. Call Service.
SV00054	Treatment Delivery Service faulted due to unexpected RTC service shutdown.	An error in the delivery system has occurred. Call Service.
SV00055	Treatment Delivery Service faulted due to unexpected RTC service disconnection.	An error in the delivery system has occurred. Call Service.
SV00056	Treatment Delivery Service faulted due to unexpected MRI Scan Service shutdown.	An error in the imaging system has occurred. Reboot the system. If the problem persists, call Service.
SV00057	Treatment Delivery Service faulted due to an unexpected MRI Scan Service disconnection.	An error in the delivery system has occurred. Call Service.
SV00058	Treatment Delivery Service faulted due to unexpected Data Processor shutdown.	An error in the delivery system has occurred. Call Service.
SV00059	Treatment Delivery Service faulted due to unexpected Data Processor disconnection.	An error in the delivery system has occurred. Call Service.
SV00060	Treatment Delivery Service faulted due to unexpected Service Manager shutdown.	An error in the delivery system has occurred. Call Service.
SV00061	Treatment Delivery Service faulted due to unexpected Service Manager disconnection.	An error in the delivery system has occurred. Call Service.
SV00062	Treatment Delivery Service Faulted. Sending gating beam on/off signal failed.	An error in the delivery system has occurred. Call Service.
SV00063	Treatment Delivery Service faulted. MRI Notification is not subscribed.	An error in the delivery system has occurred. Call Service.
SV00064	Treatment Delivery Faulted. Imaging Data notification is not subscribed.	An error in the delivery system has occurred. Call Service.
SV00065	Treatment Delivery Faulted. End Early notification is not subscribed.	An error in the delivery system has occurred. Call Service.
SV00066	Treatment Delivery Faulted. Delivery complete is not subscribed.	An error in the delivery system has occurred. Call Service.
SV00067	Treatment Delivery Faulted. Treatment status is not subscribed.	An error in the delivery system has occurred. Call Service.
SV00068	Failed processing request. Invalid MRI client.	An error in the delivery system has occurred. Call Service.
SV00069	Failed processing request. Invalid Data Processor client.	An error in the delivery system has occurred. Call Service.

ID	DIALOG MESSAGE	USER ACTION
SV00070	Failed processing request. Invalid RTC client.	An error in the delivery system has occurred. Call Service.
SV00071	Failed reading configuration file.	An error in the delivery system has occurred. Reboot the system. If the problem persists, call Service.
SV00072	Failed during pause due to an invalid pause state.	An error in the delivery system has occurred. Call Service.
SV00073	Failed during resume due to an invalid resume state.	An error in the delivery system has occurred. Call Service.
SV00074	Failed during pause due to an invalid pause state.	An error in the delivery system has occurred. Call Service.
SV00075	Failed processing treatment status data.	An error in the delivery system has occurred. Call Service.
SV00076	Service Manager failed registering a client due to invalid registration details.	An error in the delivery system has occurred. Call Service.
SV00077	Service Manager failed registering client.	An error in the delivery system has occurred. Call Service.
SV00078	Service Manager failed registering client = {0}.	An error in the delivery system has occurred. Call Service.
SV00079	Service Manager failed adding client systems.	An error in the delivery system has occurred. Call Service.
SV00080	Service Manager failed subscription for client due to invalid subscription details.	An error in the delivery system has occurred. Call Service.
SV00081	Service Manager failed subscription for client = {0}.	An error in the delivery system has occurred. Call Service.
SV00082	Service Manager failed publishing a heartbeat due to invalid hearbeat details.	An error in the delivery system has occurred. Call Service.
SV00083	Service Manager failed publishing a heartbeat for client = {0}.	An error in the delivery system has occurred. Call Service.
SV00084	Service Manager failed publishing an interlock due to invalid interlock details.	An error in the delivery system has occurred. Call Service.
SV00085	Service Manager failed publishing an interlock for client = {0}.	An error in the delivery system has occurred. Call Service.
SV00086	Service Manager failed publishing a system fault due to invalid fault details.	An error in the delivery system has occurred. Call Service.
SV00087	Service Manager failed publishing a system fault for client = {0}.	An error in the delivery system has occurred. Call Service.
SV00088	Service Manager failed requesting clear system faults.	An error in the delivery system has occurred. Call Service.
SV00089	Service Manager failed setting the subsystem state due to invalid state details.	An error in the delivery system has occurred. Call Service.

ID	DIALOG MESSAGE	USER ACTION
SV00090	Service Manager failed setting the subsystem state for client ={0}.	An error in the delivery system has occurred. Call Service.
SV00091	Service Manager failed unsubscribing for client = {0}.	An error in the delivery system has occurred. Call Service.
SV00092	Service Manager failed unregistering client = {0}.	An error in the delivery system has occurred. Call Service.
SV00093	Failed adding subscription for the client = {0} due to invalid event type.	An error in the delivery system has occurred. Call Service.
SV00094	Failed adding heartbeat subscription for the client = {0} due to no publishers specified.	An error in the delivery system has occurred. Call Service.
SV00095	Failed adding Interlock subscription for the client = {0} due to no publishers specified.	An error in the delivery system has occurred. Call Service.
SV00096	Failed adding System Fault subscription for the client = {0} due to no publishers specified.	An error in the delivery system has occurred. Call Service.
SV00097	Application has not been activated and none of the Services will be started. Call ViewRay service group to activate your application.	An error in the delivery system has occurred. Call Service.
SV00098	Services file consistency check failed and none of the Services will be started. Call ViewRay service group.	An error in the delivery system has occurred. Call Service.
SV00099	RTC Service could not start. Failed connecting to Service Manager.	An error in the delivery system has occurred. Call Service.
SV00100	RTC Service failed opening delivery session due to invalid fraction details.	An error in the delivery system has occurred. Call Service.
SV00101	RTC Service cannot process request to {0} due to {1}.	An error in the delivery system has occurred. Call Service.
SV00102	RTC Service failed opening delivery session for treatment record Id = {0}.	An error in the delivery system has occurred. Call Service.
SV00103	RTC Service failed starting delivery for treatment record Id = {0}.	An error in the delivery system has occurred. Call Service.
SV00104	RTC Service failed pausing delivery for treatment record Id = {0}.	An error in the delivery system has occurred. Call Service.
SV00105	RTC Service failed resuming delivery for treatment record Id = {0}.	An error in the delivery system has occurred. Call Service.
SV00106	RTC Service failed ending delivery early for treatment record Id = {0}.	An error in the delivery system has occurred. Press e-stop to stop delivery. Call Service.
SV00107	RTC Service failed closing delivery for treatment record Id = {0}.	An error in the delivery system has occurred. Call Service.

ID	DIALOG MESSAGE	USER ACTION
SV00108	RTC Service failed getting delivery data for treatment record Id = {0}.	An error in the delivery system has occurred. Call Service.
SV00109	RTC Service failed deleting delivery data for treatment record Id = {0}.	An error in the delivery system has occurred. Call Service.
SV00110	RTC Service failed sending {0} command to RTC.	An error in the delivery system has occurred. Call Service.
SV00111	RTC Service failed sending {0} command to RTC.	An error in the delivery system has occurred. Call Service.
SV00112	RTC Couch Service failed unexpectedly during opening a couch session	An error in the delivery system has occurred. Call Service.
SV00113	RTC Service failed unexpectedly during sending plan couch position to RTC	An error in the delivery system has occurred. Call Service.
SV00114	RTC Couch Service failed unexpectedly during closing a couch session	An error in the delivery system has occurred. Call Service.
SV00115	RTC Couch Service failed getting the couch client channel details	An error in the delivery system has occurred. Call Service.
SV00116	RTC Couch Service failed initiating open sesssion request since the service state is invalid	An error in the delivery system has occurred. Call Service.
SV00117	RTC Service failed completing processing open sesssion request since the service state is invalid	An error in the delivery system has occurred. Call Service.
SV00118	RTC Couch Service failed initiating sending position request since the service state is invalid	An error in the delivery system has occurred. Call Service.
SV00119	RTC Couch Service failed completing processing sending position request since the service state is invalid	An error in the delivery system has occurred. Call Service.
SV00120	RTC Couch Service failed initiating close session request since the service state is invalid	An error in the delivery system has occurred. Call Service.
SV00121	RTC Couch Service failed completing processing close session request since the service state is invalid	An error in the delivery system has occurred. Call Service.
SV00122	RTC Couch Service failed un registering from couch position notifications	An error in the delivery system has occurred. Call Service.
SV00123	RTC Couch Service failed registering to couch position notifications	An error in the delivery system has occurred. Call Service.
SV00124	RTC Couch Service failed sending couch position to RTC system	An error in the delivery system has occurred. Call Service.
SV00125	Failed reading configuration file.	An error in the delivery system has occurred. Call Service.

ID	DIALOG MESSAGE	USER ACTION
SV00126	RTC Service could not re-connect to RTC System, exceeded maximum retries. RTC Service must be restarted.	An error in the delivery system has occurred. Call Service.
SV00127	Failed connecting to RTC system.	An error in the delivery system has occurred. Call Service.
SV00128	Already connected to RTC System.	An error in the delivery system has occurred. Call Service.
SV00129	Socket connection to RTC system failed.	An error in the delivery system has occurred. Call Service.
SV00130	Failed closing socket connection to RTC system.	An error in the delivery system has occurred. Call Service.
SV00131	Timed out disconnecting from RTC system.	An error in the delivery system has occurred. Call Service.
SV00132	Failed disconnecting to RTC system.	An error in the delivery system has occurred. Call Service.
SV00133	Socket failure sending pause to RTC system.	An error in the delivery system has occurred. Call Service.
SV00134	Timed out receiving pause ack from RTC system.	An error in the delivery system has occurred. Call Service.
SV00135	Failed sending pause to RTC system.	An error in the delivery system has occurred. Call Service.
SV00136	Socket failure sending resume to RTC system.	An error in the delivery system has occurred. Call Service.
SV00137	Timed out receiving resume ack from RTC system.	An error in the delivery system has occurred. Call Service.
SV00138	Failed sending resume to RTC system.	An error in the delivery system has occurred. Call Service.
SV00139	Socket failure getting configuration from RTC system.	An error in the delivery system has occurred. Call Service.
SV00140	Sending plan to RTC system failed due to {0}.	An error in the delivery system has occurred. Call Service.
SV00141	Failed during sending to RTC system.	An error in the delivery system has occurred. Call Service.
SV00142	Retrieving treatement result failed due to {0}.	An error in the delivery system has occurred. Call Service.
SV00143	Failed receiving data from RTC subsystem.	An error in the delivery system has occurred. Call Service.
SV00144	Timed out receiving end early ack from RTC system.	An error in the delivery system has occurred. Call Service.
SV00145	Failed sending end early to RTC system.	An error in the delivery system has occurred. Call Service.

ID	DIALOG MESSAGE	USER ACTION
SV00146	Treatment Delivery cannot be started since Services and RTC versions are incompatible.	An error in the delivery system has occurred. Call Service.
SV00147	Treatment Delivery cannot start since communication layer failed sending Get Ready status request to RTC system.	An error in the delivery system has occurred. Call Service.
SV00148	Treatment Delivery cannot start since communication layer timed out getting an respose for the Get Ready status request from RTC system.	An error in the delivery system has occurred. Call Service.
SV00149	Treatment Delivery cannot start since RTC system is still initializing.	Wait a few minutes to start delivery. If the problem persists, call Service.
SV00150	Treatment Delivery cannot start since RTC system is still initializing and dose information for previous treatment still exists.	An error in the delivery system has occurred. Call Service.
SV00151	Treatment Delivery cannot start since RTC system is still initializing and faulted.	An error in the delivery system has occurred. Call Service.
SV00152	Treatment Delivery cannot start since RTC system is still initializing, faulted, and unretrieved dose information for previous treatment still exists.	An error in the delivery system has occurred. Call Service.
SV00153	Treatment Delivery cannot start since RTC system is faulted and unretrieved dose information for previous treatment still exists.	An error in the delivery system has occurred. Call Service.
SV00154	Treatment Delivery cannot start since RTC system is delivering another treatment.	An error in the delivery system has occurred. Call Service.
SV00155	Treatment Delivery cannot start since RTC system is faulted.	An error in the delivery system has occurred. Call Service.
SV00156	Treatment Delivery cannot start since unretrieved dose information for previous treatment still exists.	An error in the delivery system has occurred. Call Service.
SV00157	Treatment Delivery cannot start since RTC system is having an unknown techincal problem. Contact Service.	An error in the delivery system has occurred. Call Service.
SV00158	Failed sending a Start delivery command to RTC system.	An error in the delivery system has occurred. Call Service.
SV00159	Failed sending End Early command due to invalid value.	An error in the delivery system has occurred. Call Service.
SV00160	Failed sending End Early command to RTC system.	An error in the delivery system has occurred. If delivery is proceeding, pre e-stop. Call Service.
SV00161	Failed seding Delete command to RTC system.	An error in the delivery system has occurred. Call Service.

ID	DIALOG MESSAGE	USER ACTION
SV00162	Failed sending command to RTC system. Invalid/Unknown command received	An error in the delivery system has occurred. Call Service.
SV00163	Failed sending command to RTC system.	An error in the delivery system has occurred. Call Service.
SV00164	Data Processor Service could not be started due to failure connecting to Service Manager.	An error in the delivery system has occurred. Call Service.
SV00165	Data Processor service failed processing request due to invalid data: {0}.	An error in the delivery system has occurred. Call Service.
SV00166	Data Processor service cannot process request to due to {0}.	An error in the delivery system has occurred. Call Service.
SV00167	Data Processor service failed opening processing session.	An error in the delivery system has occurred. Call Service.
SV00168	Data Processor service failed closing the session.	An error in the delivery system has occurred. Call Service.
SV00169	Data Processor service failed adding work processors.	An error in the delivery system has occurred. Call Service.
SV00170	Data Processor service failed adding data image item to work processor.	An error in the delivery system has occurred. Call Service.
SV00171	Data Processor service failed removing the work item.	An error in the delivery system has occurred. Call Service.
SV00172	Data Processor service failed starting the work processor.	An error in the delivery system has occurred. Call Service.
SV00173	Data Processor service failed stopping the work processor.	An error in the delivery system has occurred. Call Service.
SV00174	Data Processor service failed due to invalid scan image data.	An error in the delivery system has occurred. Call Service.
SV00175	Beam Off detected consecutively for the gating terminate wait time.	Reposition the patient. Attempt to restart delivery and try again. If the problem persists, call Service.
SV00176	Data Processor service failed processing due to algorithm engine initialization failure.	An error in the delivery system has occurred. Call Service.
SV00177	Data Processor service failed processing due to actual versus input image plane mismatch.	An error in the imaging system has occurred. Call Service.
SV00178	Data Processor service failed processing due to gating setup failure	An error in the delivery system has occurred. Call Service.
SV00179	Data Processor service failed processing due to insufficient overlap on gating region.	The gating setup appears incorrect Setup gating planes, target, and boundary again and retry. If the problem persists, call Service.
SV00180	Data Processor service failed processing due to failure of the gating process.	An error in the delivery system has occurred. Call Service.

ID	DIALOG MESSAGE	USER ACTION
SV00181	Data Processor service failed processing due to notification processor time out receving image data.	An error in the delivery system has occurred. Call Service.
SV00182	Data Processor service failed processing due to storage processor time out receving image data.	An error in the imaging system has occurred. Call Service.
SV00183	Image Data Change notification is unsubscribed.	An error in the delivery system has occurred. Call Service.
SV00184	Gating decision notification is unsubscribed.	An error in the delivery system has occurred. Call Service.
SV00185	Lost communication with MRI Scan Service. Restart the system, call service if problem persists.	An imaging system error has occurred. Reboot the system. If the problem persists, call service.
SV00186	MRI Scan service failed processing request because an invalid type was recevied.	An imaging system error has occurred. Reboot the system. If the problem persists, call service.
SV00187	MRI Scan service failed opening scan session.	An imaging system error has occurred. Reboot the system. If the problem persists, call service.
SV00188	MRI Scan service failed registering patient.	An imaging system error has occurred. Reboot the system. If the problem persists, call service.
SV00189	MRI Scan service failed getting scan protocol.	An imaging system error has occurred. Reboot the system. If the problem persists, call service.
SV00190	MRI Scan service failed getting last scan result.	An imaging system error has occurred. Reboot the system. If the problem persists, call service.
SV00191	MRI Scan service failed getting scan protocol info.	An imaging system error has occurred. Reboot the system. If the problem persists, call service.
SV00192	MRI Scan service failed getting scan protocol tree node list.	An imaging system error has occurred. Reboot the system. If the problem persists, call service.
SV00193	MRI Scan service failed closing patient.	An imaging system error has occurred. Reboot the system. If the problem persists, call service.
SV00194	MRI Scan service failed closing session.	An imaging system error has occurred. Reboot the system. If the problem persists, call service.
SV00195	MRI Scan service failed starting scan.	An imaging system error has occurred. Reboot the system. If the problem persists, call service.
SV00196	MRI Scan service failed stopping scan.	An imaging system error has occurred. Reboot the system. If the problem persists, call service.
SV00197	MRI Scan service failed reading configurations.	An imaging system error has occurred. Reboot the system. If the problem persists, call service.
SV00198	Send External Protocol failed starting scan.	An imaging system error has occurred. Reboot the system. If the problem persists, call service.
SV00199	MRI Scan service failed reconnecting to MRI Scanner system.	An imaging system error has occurred. Reboot the system. If the problem persists, call service.
SV00200	MRI Scan service failed connecting to MRI Scanner.	An imaging system error has occurred. Reboot the system. If the problem persists, call service.

ID	DIALOG MESSAGE	USER ACTION
SV00201	MRI Scan service failed connecting to MRI Scanner.	An imaging system error has occurred. Reboot the system. If the problem persists, call service.
SV00202	MRI Scan service failed processing request, invalid {0}.	An imaging system error has occurred. Reboot the system. If the problem persists, call service.
SV00203	MRI Scan service failed processing due to unknown type recevied.	An imaging system error has occurred. Reboot the system. If the problem persists, call service.
SV00204	MRI Scan service cannot process request to {0} due to {1}.	An imaging system error has occurred. Reboot the system. If the problem persists, call service.
SV00205	MRI Scan service failed getting the protocol info due to invalid protocol info.	An imaging system error has occurred. Reboot the system. If the problem persists, call service.
SV00206	MRI Scan service failed reading configurations.	An imaging system error has occurred. Reboot the system. If the problem persists, call service.
SV00207	Simulation provider failed reading configuration.	An imaging system error has occurred. Reboot the system. If the problem persists, call service.
SV00208	Mri Console channel has timed out.	An imaging system error has occurred. Reboot the system. If the problem persists, call service.
SV00209	Mri Console control channel connect failure.	An imaging system error has occurred. Reboot the system. If the problem persists, call service.
SV00210	Mri Console Control Channel connect exception.	An imaging system error has occurred. Reboot the system. If the problem persists, call service.
SV00211	Mri Console Control Channel disconnect failure.	An imaging system error has occurred. Reboot the system. If the problem persists, call service.
SV00212	Mri Console Control Channel disconnect exception.	An imaging system error has occurred. Reboot the system. If the problem persists, call service.
SV00213	Mri Console Control Channel registering patient failed.	An imaging system error has occurred. Reboot the system. If the problem persists, call service.
SV00214	Mri Console Control Channel registering patient failed.	An imaging system error has occurred. Reboot the system. If the problem persists, call service.
SV00215	Mri Console Control Channel closing patient failed.	An imaging system error has occurred. Reboot the system. If the problem persists, call service.
SV00216	Mri Console Control Channel closing patient failed.	An imaging system error has occurred. Reboot the system. If the problem persists, call service.
SV00217	Mri Console Control Channel starting scan failed.	An imaging system error has occurred. Reboot the system. If the problem persists, call service.
SV00218	Mri Console Control Channel starting scan failed.	An imaging system error has occurred. Reboot the system. If the problem persists, call service.
SV00219	Mri Console Control Channel stopping scan failed.	An imaging system error has occurred. Reboot the system. If the problem persists, call service.
SV00220	Mri Console Control Channel stopping scan failed.	An imaging system error has occurred. Reboot the system. If the problem persists, call service.
SV00221	Mri Console Control Channel get scan protocol tree node list failed.	An imaging system error has occurred. Reboot the system. If the problem persists, call service.

ID	DIALOG MESSAGE	USER ACTION
SV00222	Mri Console Control Channel get scan protocol tree node list failed.	An imaging system error has occurred. Reboot the system. If the problem persists, call service.
SV00223	Mri Console Control Channel get protocol info failed.	An imaging system error has occurred. Reboot the system. If the problem persists, call service.
SV00224	Mri Console Control Channel get protocol info failed.	An imaging system error has occurred. Reboot the system. If the problem persists, call service.
SV00225	Mri Console Data Channel starting server failure.	An imaging system error has occurred. Reboot the system. If the problem persists, call service.
SV00226	Mri Console Data Channel stopping server failure.	An imaging system error has occurred. Reboot the system. If the problem persists, call service.
SV00227	Mri Console Data Channel waiting for time out.	An imaging system error has occurred. Reboot the system. If the problem persists, call service.
SV00228	Mri Console Data Channel timed out while connecting.	An imaging system error has occurred. Reboot the system. If the problem persists, call service.
SV00229	Mri Console Data Channel disconnect failure.	An imaging system error has occurred. Reboot the system. If the problem persists, call service.
SV00230	Siemens remote control client connect failure.	An imaging system error has occurred. Reboot the system. If the problem persists, call service.
SV00231	Siemens remote control client connect warning.	An imaging system error has occurred. Reboot the system. If the problem persists, call service.
SV00232	Siemens remote control client connect warning.	An imaging system error has occurred. Reboot the system. If the problem persists, call service.
SV00233	Siemens remote control client disconnect failure.	An imaging system error has occurred. Reboot the system. If the problem persists, call service.
SV00234	Siemens remote control client disconnect warning.	An imaging system error has occurred. Reboot the system. If the problem persists, call service.
SV00235	Siemens remote control client disconnect info.	An imaging system error has occurred. Reboot the system. If the problem persists, call service.
SV00236	Siemens remote control client register patient failure.	An imaging system error has occurred. Reboot the system. If the problem persists, call service.
SV00237	Siemens remote control client register patient info.	An imaging system error has occurred. Reboot the system. If the problem persists, call service.
SV00238	Siemens remote control client register patient warning.	An imaging system error has occurred. Reboot the system. If the problem persists, call service.
SV00239	Siemens remote control client close patient failure.	An imaging system error has occurred. Reboot the system. If the problem persists, call service.
SV00240	Siemens remote control client close patient info.	An imaging system error has occurred. Reboot the system. If the problem persists, call service.
SV00241	Siemens remote control client close patient warning.	An imaging system error has occurred. Reboot the system. If the problem persists, call service.
SV00242	Siemens remote control client start scan failure.	An imaging system error has occurred. Reboot the system. If the problem persists, call service.

ID	DIALOG MESSAGE	USER ACTION
SV00243	Siemens remote control client start scan info.	Informational message.
SV00244	Siemens remote control client start scan warning.	An imaging system error has occurred. Reboot the system. If the problem persists, call service.
SV00245	Siemens remote control client stop scan failure.	An imaging system error has occurred. Reboot the system. If the problem persists, call service.
SV00246	Siemens remote control client stop scan info.	Informational message.
SV00247	Siemens remote control client stop scan warning.	Informational message.
SV00248	Siemens remote control client GetProtocolInfo failure.	An imaging system error has occurred. Reboot the system. If the problem persists, call service.
SV00249	Siemens remote control client GetProtocolInfo info.	Informational message.
SV00250	Siemens remote control client GetProtocolInfo warning.	Informational message.
SV00251	Siemens remote control client GetProtocolTreeNode failure.	An imaging system error has occurred. Reboot the system. If the problem persists, call service.
SV00252	Siemens remote control client GetProtocolTreeNode info.	Informational message.
SV00253	Siemens remote control client GetProtocolTreeNode warning.	Informational message.
SV00254	Console failed connecting to Service Manager.	An error in the delivery system has occurred. Call Service.
SV00255	Console exceeded maximum retries connecting to Service Manager. Console must be restarted.	An error in the delivery system has occurred. Reboot the system. If the problem persists, call Service.
SV00256	Console failed instantiating Service Manager Client.	An error in the delivery system has occurred. Call Service.
SV00257	Console is disconnected from Service Manager, will attempt to reconnect in the background.	An error in the delivery system has occurred. Reboot the system. If the problem persists, call Service.
SV00258	Console has lost communication with Service Manager, will attempt to reconnect in the background.	An error in the delivery system has occurred. Reboot the system. If the problem persists, call Service.
SV00259	This Console Failed to Connect to the Service Computer. Please Click OK and 3 additional attempts to connect will be made. This will take approximately 2 minutes. If a connection cannot be established, the Console will function only as a Treatment Planning System.	An error in the delivery system has occurred. Reboot the system. If the problem persists, call Service.

ID	DIALOG MESSAGE	USER ACTION
SV00260	Console failed instantiating Treatment Delivery Service Client.	An error in the delivery system has occurred. Reboot the system. If the problem persists, call Service.
SV00261	Console failed connecting to Treatment Delivery Service.	An error in the delivery system has occurred. Reboot the system. If the problem persists, call Service.
SV00262	Console failed to {0} due to invalid treatment client.	An error in the delivery system has occurred. Call Service.
SV00263	Treatment Delivery Service Client failed Starting the Delivery.	An error in the delivery system has occurred. Call Service.
SV00264	Treatment Delivery Service Client failed Pausing the Delivery.	An error in the delivery system has occurred. Call Service.
SV00265	Treatment Delivery Service Client failed Resuming the Delivery.	An error in the delivery system has occurred. Call Service.
SV00266	Treatment Delivery Service Client failed Stopping the Delivery.	An error in the delivery system has occurred. Call Service.
SV00267	Treatment Delivery Service Client failed Closing the Delivery.	An error in the delivery system has occurred. Call Service.
SV00268	Treatment Delivery Service Client failed getting the treatment data.	An error in the delivery system has occurred. Call Service.
SV00269	Treatment Delivery Service Client failed sending delete treatment data signal.	An error in the delivery system has occurred. Call Service.
SV00270	Begin Treatment failed due to invalid Treatment Delivery protocol.	An error in the delivery system has occurred. Call Service.
SV00271	Failed during reading the configuration file.	An error in the delivery system has occurred. Call Service.
SV00272	Failed connecting to Service Manager, invalid service manager client.	An error in the delivery system has occurred. Call Service.
SV00273	Failed connecting to Service Manager.	An error in the delivery system has occurred. Call Service.
SV00274	Instanstiating MRI Scan Service Client failed.	An error in the imaging system has occurred. Reboot the system. If the problem persists, call Service.
SV00275	MRI Scan Service client failed connecting to MRI Scan Service.	An error in the imaging system has occurred. Reboot the system. If the problem persists, call Service.
SV00276	MRI Scan Service Client failed opening session.	An error in the imaging system has occurred. Reboot the system. If the problem persists, call Service.
SV00277	MRI Scan Service Client failed starting scan.	An error in the imaging system has occurred. Reboot the system. If the problem persists, call Service.

ID	DIALOG MESSAGE	USER ACTION
SV00278	MRI Scan Service Client failed getting the protocol information.	An error in the imaging system has occurred. Reboot the system. If the problem persists, call Service.
SV00279	MRI Scan Service Client failed getting the protocol tree node information.	An error in the imaging system has occurred. Reboot the system. If the problem persists, call Service.
SV00280	MRI Scan Service Client failed stopping scan.	An error in the imaging system has occurred. Reboot the system. If the problem persists, call Service.
SV00281	MRI Scan Service Client failed closing session.	An error in the imaging system has occurred. Reboot the system. If the problem persists, call Service.
SV00282	MRI Scan Service Client failed processing request due to invalid connection state.	An error in the imaging system has occurred. Reboot the system. If the problem persists, call Service.
SV00283	Data Processor Service Client failed processing request due to invalid connection state.	An error in the delivery system has occurred. Reboot the system. If the problem persists, call Service.
SV00284	Instanstiating Data Processor Service Client failed.	An error in the delivery system has occurred. Reboot the system. If the problem persists, call Service.
SV00285	Data Processor Service Client failed connecting to Data Processor Service.	An error in the delivery system has occurred. Reboot the system. If the problem persists, call Service.
SV00286	Data Processor Service Client failed opening session.	An error in the delivery system has occurred. Reboot the system. If the problem persists, call Service.
SV00287	Data Processor Service Client failed closing session.	An error in the delivery system has occurred. Reboot the system. If the problem persists, call Service.
SV00288	RTC Service Client failed connecting to RTC Service.	An error in the delivery system has occurred. Call Service.
SV00289	RTC Service Client failed processing request due to invalid connection state.	An error in the delivery system has occurred. Call Service.
SV00290	RTC Service Client failed opening delivery session.	An error in the delivery system has occurred. Call Service.
SV00291	RTC Service Client failed starting a delivery session.	An error in the delivery system has occurred. Call Service.
SV00292	RTC Service Client failed pausing a delviery session.	An error in the delivery system has occurred. Call Service.
SV00293	RTC Service Client failed resuming a delivery session.	An error in the delivery system has occurred. Call Service.
SV00294	RTC Service Client failed stopping delivery.	An error in the delivery system has occurred. Call Service.

ID	DIALOG MESSAGE	USER ACTION
SV00295	RTC Service Client failed closing delivery session.	An error in the delivery system has occurred. Call Service.
SV00296	RTC Service Client failed retrieving the treatment data.	An error in the delivery system has occurred. Call Service.
SV00297	RTC Service Client failed sending acknowledgement on data recevied.	An error in the delivery system has occurred. Call Service.
SV00298	RTC Service Client failed sending gating beam on/off command.	An error in the delivery system has occurred. Call Service.
SV00299	RTC Service Couch Client failed connecting to RTC Service Couch	An error in the delivery system has occurred. Call Service.
SV00300	RTC Service Couch Client failed opening the couch session	An error in the delivery system has occurred. Call Service.
SV00301	RTC Service Couch Client failed closing the couch session	An error in the delivery system has occurred. Call Service.
SV00302	RTC Service Couch Client failed sending the couch position	An error in the delivery system has occurred. Call Service.
SV00303	Opening Service Manager failed updating service properties.	An error in the delivery system has occurred. Call Service.
SV00304	Opening Service failed attempting to update the service recovey properties.	An error in the delivery system has occurred. Call Service.
SV00305	Failed while locking the Service for updating properties	An error in the delivery system has occurred. Call Service.
SV00306	Access was denied during an attempt to update the service properties.	An error in the delivery system has occurred. Call Service.
SV00307	Setting the Recovery on Failure properties for service failed.	An error in the delivery system has occurred. Call Service.
SV00308	Failed setting the recovery on failure properties for the service.	An error in the delivery system has occurred. Call Service.
SV00309	Data Processor service has faulted. Gating data processor timed out receving image data.	An error in the delivery system has occurred. Call Service.
SV00310	MRI Scan service was has faulted due to invalid data.	An error in the imaging system has occurred. Call Service.
AE00001	An unknown error has occurred.	Reboot the system and report the error to service.
AE00002	A bad parameter was detected.	Check the data. If the problem persists, call service.
AE00003	All dimensions of volume must be greater than one. Single slice volumes are invalid.	Use another volume for processing or import the entire volume.
AE00004	A calculation error has occurred.	Try the operation again. Reboot the system and try again. Check the data. If the problem persists, call service.

ID	DIALOG MESSAGE	USER ACTION
AE00005	An exception has occurred during algorithm calculation.	Reboot the system and try again. If the problem persists, call service.
AE00006	A memory allocation error has occurred.	Reboot the system and try again. If the problem persists, call service.
AE00007	A memory deallocation error has occurred.	Reboot the system and try again. If the problem persists, call service.
AE00008	Conformal beam with closed multileaf collimator detected. Reset any closed conformal beams to have multileaf collimator openings, or turn the beams off.	Adjust or disable the beam.
AE00009	Maximum margin is {0}.	Choose a lower value that is in rang.
AE00010	An IMRT beam does not have valid segments. One or more segments are fully closed or have non-positive fluence values.	Call service.
AE00011	Conformal beam with the multileaf collimator not open wide enough has been detected. Reset any small conformal beams to have larger multileaf collimator openings, or turn the beams off.	Adjust or disable beam.
AE00012	Not enough image overlap to compute registration.	Choose two volumes for registration that overlap.
AE00013	Error reading configuration file for {0}.	Reboot the system and try again. If the problem persists, call service.
AE00014	Error reading {0} configuration entry {1}.	Reboot the system and try again. If the problem persists, call service.
AE00015	The structure contains no voxels. Draw contours or remove the structure.	Draw contours or remove the structure.
AE00016	Operation cannot be performed on modality of type {0}.	Use only CT, PET, and MRI image sets in the system.
AE00017	PET series with units {0} cannot be processed.	Delete the series from the system. Re-import the series. Try the operation again.
AE00018	Couch x axis must be increasing and y axis must be decreasing to complete processing.	Call service.
AE00019	The current couch placement has resulted in an empty feasible region. You cannot place any valid isocenters until the couch position is adjusted.	Adjust couch. Use auto couch placement tools.
AE00020	One or more MLCs are interdigitated. Adjust leaf positions to remove interdigitation.	Adjust the leaf positions to remove interdigitation and retry.

ID	DIALOG MESSAGE	USER ACTION
AE00021	Some beams may cover the non-treatable region through couch ends, please re-adjust couch position and/or beam isocenter location.	Readjust the couch or the beam isocenter location.
AE00022	Number of levels out of range. Min={0}. Max={1}.	Select a value for the levels that is in the acceptable range.
AE00023	Dose resolution is out of range. Min={0}. Max={1}.	Choose a dose resoution value that is in the valid range.
AE00024	Tracking is enabled but there are no tracking points.	Checking gating plane setup. If the problem persists, call service.
AE00025	More tracking points than result points.	Call service.
AE00026	Patient treatment date is earlier than source calibration date.	Call service.
AE00027	Beam-on time conversion failed.	Call service.
AE00028	Fixed conformal beam openings at angles {0} do not contain the beam's weight point.	Change the beam angle or weight point such that the beam angle contains the weight point.
VZ00001	Unknown Visualization error.	An error has occurred inthe visualization subsystem. Reboot the system. If the problem persists, call service.
VZ00002	A bad parameter was detected in the Visualization system. '{0}'	Try the operation again. Reboot the system. If the problem persists, call service.
VZ00003	An exception has occurred in the Visualization system.	An unknown error has occurred in the visualization subsystem. Reboot the system. If the problem persists, call service.
VZ00004	A memory allocation error has occurred in the Visualization system.	An unknown error has occurred in the visualization subsystem. Reboot the system. If the problem persists, call service.
VZ00005	An error in image rendering has occurred.	An unknown error has occurred in the visualization subsystem. Reboot the system. If the problem persists, call service.
VZ00006	Patient orientation and direction are not consistent for both volumes. '{0}'	Delete the data from the system. Re-import the data. If the problem persists, call Service.
UI00001	An exception occured on the Console.	An error in the system has occurred. Reboot the system. If the problem persists, call Service.
UI00002	The RTC plan has not been validated.	An error in the delivery system has occurred. Reboot the system. If the problem persists, call Service.
UI00003	The gating plane visualization data is not valid.	An error in the delivery system has occurred. Reboot the system. If the problem persists, call Service.
UI00004	Invalid beam segment returned.	An error in the delivery system has occurred. Reboot the system. If the problem persists, call Service.

ID	DIALOG MESSAGE	USER ACTION
UI00005	Could not send the Treatment Delivery protocol to Services.	An error in the delivery system has occurred. Reboot the system. If the problem persists, call Service.
UI00006	Could not start the heartbeat to monitor Services.	An error in the delivery system has occurred. Reboot the system. If the problem persists, call Service.
UI00007	Configuration {0} not found. The system will use default configuration.	The user should first try again. If the problem persists, try restarting. If the problem still persists, call service.
UI00008	The Console uninterruptable power supply (UPS) state is not available. Check the UPS USB cable and/or call service. Treatment delivery is in progress and will be ended in {0} minutes at {1}. The Console will be shut down automatically in {2} minutes at {3}. If return to AC power is detected, shutdown will not occur.	The user should first ensure the UPS USB cable is securely connected to both the UPS and the machine. If so, then the user may optionally end treatment early before the system auto-ends, or wait for the system to end treatment and shut down.
UI00009	The Console uninterruptable power supply (UPS) state is not available. Check the UPS USB cable and/or call service. The Console will be shut down automatically in {0} minutes at {1}. If return to AC power is detected, shutdown will not occur.	The user should first ensure the UPS USB cable is securely connected to both the UPS and the machine. If so, then the user may optionally shut down.
UI00010	The Services system uninterruptable power supply (UPS) state is not available. Check the UPS USB cable and/or call service. If treatment delivery is in progress, it will be ended in {0} minutes. If return to AC power is detected, treatment delivery will not be stopped.	The user should first ensure the UPS USB cable is securely connected to both the UPS and the machine. If so, then the user may optionally shut down.
UI00011	The console is now running on AC power supply.	No user action is required.
UI00012	The services computer is now running on AC power supply.	No user action is required.
UI00013	The Console has detected that it is running on battery. Check the UPS power cables and/or call service. Treatment delivery is in progress and will be ended in {0} minutes at {1}. The Console will be shut down automatically in {2} minutes at {3}. If return to AC power is detected, shutdown will not occur.	The user should check the power source and plug if the building power supply is not affected. The user may optionally end treatment and shut down before the system does so automatically.

ID	DIALOG MESSAGE	USER ACTION
UI00014	The Console has detected that it is running on battery. Check the UPS power cables and/or call service. The console will be shut down automatically in {0} minutes at {1}. If return to AC power is detected, shutdown will not occur.	The user should check the power source and plug if the building power supply is not affected. The user may optionally shut down before the system does so automatically.
UI00015	The Services computer has detected that it is running on battery. Check the UPS power cables and/or call service. If treatment delivery is in progress, it will be ended in {0} minutes at {1} and services system will be shutdown automatically. If return to AC power is detected, treatment delivery will not be stopped.	The user should check the power source and plug if the building power supply is not affected. The user may optionally end treatment and shut down before the system does so automatically.
UI00016	This application has not been activated and will be shut down. Activation requires a valid license. Call service.	Call service to activate the software.
UI00017	This application did not pass the corruption check and will be shutdown. Call service.	Call service to check the files and reset the CRC.
UI00018	Error loading beam template.	The user should try to load the template again, and check the connection to the database.
UI00019	Error saving beam template.	The user should try to save the template again, and check the connection to the database.
UI00020	Contour Smoothing Error	The user may try to smooth the contours again using modified settings.
UI00021	Algorithm engine status is returned inactive, pending, or active on callback. Retry.	Try operation again. If problem persists, reboot the system and try again. If problem still persists, call service
UI00022	An error occured while filling contour holes.	The user may adjust parameters and try again.
UI00023	An error occured while removing contour islands.	The user may adjust parameters and try again.
UI00025	The system cannot find the correct structure for operating the normalize by primary prescription operation. Check that the primary prescription structure exists in the plan.	The user must add or resolve naming issues in the plan to match the prescription's primary reference and try again.
UI00026	The system cannot find the correct point for operating the normalize by primary prescription operation. Check that the primary prescription point exists in the plan.	The user must add or resolve naming issues in the plan to match the prescription's primary reference and try again.

ID	DIALOG MESSAGE	USER ACTION
UI00027	The prescription constraint structure was not found in the plan.	The user must add or resolve naming issues in the plan to match the prescription constraint's structure and try again.
UI00028	Structure statistics must be computed to normalize to this statistic.	The user must reoptimize dose to regenerate a dose volume and statistics, and then may retry.
UI00029	The prescription constraint point was not found in the plan.	The user must add or resolve naming issues in the plan to match the prescription constraint's point and try again.
UI00030	By leaving this workflow you will lose any unsaved data.	The user may select OK to continue and discard changes, or may cancel to stay on the current screen without discarding the plan.
UI00031	PDF creation unsuccessful. Check that the file is not open or otherwise in use, and try again.	The user may try saving again with a different filename, or they may close the open PDF file and save to overwrite.
UI00032	Could not retrieve Radiation Oncologist from database.	The user should check the connection cables and try again. If the problem persists, call service.
UI00033	Could not retrieve Medical Physicist from database.	The user should check the connection cables and try again. If the problem persists, call service.
UI00034	Could not retrieve Dosimetrist from database.	The user should check the connection cables and try again. If the problem persists, call service.
UI00035	You must provide a plan name.	The user should enter a plan name and try again.
UI00036	The plan was not saved. Could not create the new plan.	The user should check the connection cables and try again. If the problem persists, call service.
UI00037	The plan was not saved. Could not check if the plan name already exists.	The user should check the connection cables and try again. If the problem persists, call service.
UI00038	The plan name "{0}" already exists for this patient. Enter a different plan name and try again.	The user should enter a different plan name and try again.
UI00039	There was an error saving the plan.	The user should check the connection cables and try again. If the problem persists, call service.
UI00040	Warning: Patient identifier mismatch. ({0})	The user may correct the data and re-import if desired.
UI00041	Error: File is for a different patient. Cannot continue. ({0})	The user may correct the data to be imported on another system, and try again. Alternatively, the user may open the correct patient for import.
UI00042	Error: File does not exist. ({0})	The user may restore the file manually and retry.
UI00043	Error: File is not a valid DICOM file. ({0})	The user may re-export the file(s) in a valid DICOM format on the originating system and retry.
UI00044	Error: File contains invalid data. Check the log for more information ({0})	The user may optionally call service to check the logs if they detect a problem with the file after opening it in the ViewRay System.
UI00045	Error: File has already been imported. ({0})	The user may find the files which already exist in the system to use.

ID	DIALOG MESSAGE	USER ACTION
UI00046	Error: Unsupported Modality. ({0})	The user may not import unsupported modality images.
UI00047	Error: Cannot import dose files that do not contain grid-based data. ({0})	Ther user may not import unsupported dose data types.
UI00048	Error: Image is not 16-bit. Only 16-bit images are supported. ({0})	Ther user may not import unsupported image data types.
UI00049	Could not retrieve the stored AE list.	Provide connection information or close the PACS import window and try again.
UI00050	Could not retrieve the stored AE information.	Provide connection information or close the PACS import window and try again.
UI00051	Could not add DICOM AE settings.	Provide connection information or close the PACS import window and try again.
UI00052	Could not delete AE from repository.	Try deleting again. If problem persists, call service.
UI00053	Could not submit transfer job.	Try again. If problem persists, call service
UI00054	An error occurred either trying to create or find the output directory. Verify output location format.	Select existing dirctory you have user rights to and try again.
UI00055	Could not start export.	If other errors are listed, correct them and try again. If problem persists call service.
UI00056	Could not export plan.	Correct any other issues listed and try to export again. If the problem persists, call service.
UI00057	There was an error querying the PACS.	Check the connection information, and ensure that the PACS is available then try again.
UI00058	Invalid Provider Selection	Select an import provider and try again. If the problem persists, call service.
UI00060	Error importing structures and contours.	The user may try exporting the structures/contours from the originating system again, and retry. The system cannot import unsupported formats.
UI00061	Error importing points.	The user may try exporting the point from the originating system again, and retry. The system cannot import unsupported formats.
UI00062	Could not add diagnosis.	The user should check the connection to the database, and retry. If the problem persists, call service.
UI00063	Could not add patient history record.	The user should check the connection to the database. If the problem persists, call service.
UI00064	Could not add patient.	The user should check the connection to the database, and retry. If the problem persists, call service.
UI00065	Could not authenticate user.	The user should try again, checking the username and password. If the error continues, contact Service.

ID	DIALOG MESSAGE	USER ACTION
UI00066	There was an error attempting to authenticate user.	The user should check the connection to the database, and the entered username and password, and retry. If the problem persists, call service.
UI00067	There was an error retrieving the permissions for the user.	The user should check the connection to the database, and retry. If the problem persists, call service.
UI00068	Unable to record into patient history.	The user should check the connection to the database. If the problem persists, call service.
UI00069	Authentication Failed	The user may retry entering their username/ password to authenticate. The site system administrator may be called to reset their password if the user forgot their password.
UI00070	Username is required. Enter a username and try again.	Enter a username and retry.
UI00071	Old password is required. Enter your password and try again.	Enter the correct password and try again.
UI00072	Password is required. Enter a password and try again.	Enter a new password in both textboxes and try again.
UI00073	Entered password values do not match. Re-enter password values and try again.	Enter the same new password in both textboxes and try again.
UI00074	An error occured while getting the delivered plan.	Restart the system. If the error persists, call service.
UI00075	A preprocessing error occured when getting the delivered plan.	Restart the system. If the error persists, call service.
UI00076	A preprocessing error occured when calculating the remaining plan.	Restart the system. If the error persists, call service.
UI00077	Could not retrieve couch model information.	The user should check the connection to the database, and retry. If the problem persists, call service.
UI00078	There was an error in the algorithm engine.	The user may check the inputs, and try again. If the problem persists, call service.
UI00079	There was an error in the algorithm engine. The error code is unknown.	The user may check the inputs, and try again. If the problem persists, call service.
UI00080	Statistics computation failed for structures:	The user should delete structures with zero volume, or add contours so that a volume is present. The user may retry optimizing or calculating dose. If the problem persists, call service.
UI00081	. Only structures with valid results will be displayed. This error may be caused by a structure with zero volume.	The user should delete structures with zero volume, or add contours so that a volume is present. The user may retry optimizing or calculating dose. If the problem persists, call service.

ID	DIALOG MESSAGE	USER ACTION
UI00082	An error occured while recalculating the DVH plot.	The user may adjust parameters and retry calculating dose. If the problem persists, call service.
UI00083	The returned DVH structure cannot be selected in the plan.	Try again. If the problem persists, reboot the system. If the problem still persists, call service.
UI00084	Invalid DVH returned from algorithm engine.	Try again. If the problem persists, reboot and try again. If the problem still persists, call service.
UI00085	The dose at a point of interest could not be calculated.	The user should check that the point resides inside the patient or a contour. The user may try calculating dose again. If the problem persists, call service.
UI00086	Could not load CT to electron density calibration.	The user should check the connection to the database, and retry. If the problem persists, call service.
UI00087	Electron density map creation algorithmic error.	The user should check the Electron Density settings for the plan and retry. If the problem persists, call service.
UI00088	PDF creation unsuccessful. Check that the file is not open or otherwise in use, and try again.	The user may try saving again with a different filename, or they may close the open PDF file and save to overwrite.
UI00089	Error calculating the feasible region for isocenters.	The user should adjust couch location to a reasonable location, if not already done. If the problem persists for normal couch-patient setups, call service.
UI00090	The {0} were modified since dose was last optimized. If you continue, the {0} will be reverted to those used to optimize dose.	Make a selection
UI00091	There was an internal error determining what should be reverted.	Nothing
UI00092	There was an error saving the plan.	The user should check the connection to the database, and retry. If the problem persists, call service.
UI00093	Error adding previously delivered dose volumes. The calculated dose volume will not include previously delivered dose.	The user may re-add or re-register the previously delivered dose in the plan, and try again. If the problem persists, call service.
UI00094	The TG51 measurement date is earlier than source calibration date by '{0}' days.	The user may fix the source entry inputs, and try to save again.
UI00095	Value out of range. TG51 measurement dose to water is '{0}'. Expected between '{1}' and '{2}'.	The user may override to save, or they may fix the source entry inputs, and try to save again.
UI00096	Cannot find a matching prescription constraint in the comparison plan. The comparison plan may not have a structure or point from the prescription.	The user may add the missing structure/point to the comparison plan, and retry, or the user may view the comparison with a missing constraint.

ID	DIALOG MESSAGE	USER ACTION
UI00097	Dose normalization parameters or user inputs are not valid.	The user may adjust inputs and retry.
UI00098	Could not add diagnosis.	The user may retry. If the problem persists, call service.
UI00099	Could not save patient information.	The user may retry. If the problem persists, call service.
UI00100	The system will restart after you click OK.	After the user clicked OK, the user will observe shutdown process and then system can be restarted.
UI00101	Error.	Nothing.
UI00102	There was an error requiring shutdown, but the error description key was not found.	Call service.
UI00103	Application will be shutdown. The system heartbeat in {1} lost connection during {0} counts.	Restart. If the problem persists, call service.
UI00104	Error while centering slices on structure or point.	Retry, or the user may scroll manually or use the MPR tool to change display slice number.
UI00105	Could not retrieve user list.	The user should check the connection to the database, and retry. If the problem persists, call service.
UI00106	Could not generate a unique id.	The user may manually enter an ID value.
UI00107	The file could not be read. Please select a jpeg image.	Select a different file for import.
UI00108	Invalid type '{0}'	
UI00109	Removed {0} which caused removal of {1}.	Nothing.
UI00110	The system hardware status is unknown.	Wait for the connection to be re-established. If a connection is not found, restart software and hardware components as necessary.
UI00111	At least one structure must be selected.	Select at least one structure for bulk density calculations and retry.
UI00112	All custom densities must have a value.	Enter the missing density values and retry.
UI00113	A CT volume must be selected from the dropdown.	Select a CT volume and retry.
UI00114	Could not convert the error message to be shown.	Nothing.
UI00115	Could not search series.	Retry. If the problem persists, contact service.
UI00116	Could not delete the selected series.	If the series is in use in a plan, the user may delete those plans and then delete the series.
UI00117	MRI Scanning Error:	Follow the directions in the specific displayed error.
UI00118	MRI Scanning Exception	Retry. If the problem persists, call service.

ID	DIALOG MESSAGE	USER ACTION
UI00119	Timeout waiting for MRI acquisition to complete.	Check the setup if using a navigator scan, and retry.
UI00120	MRI Scanner returned empty volumes.	Check the parameters used for imaging and retry. If the problem persists, call service.
UI00121	MRI Scanner failed returing volumes.	Check the parameters used for imaging and retry. If the problem persists, call service.
UI00122	MRI Scanner Console has a pop up window. Please check the scanner system.	View the scanner system and take any actions required by its popup window. Close the popup window to continue using the ViewRay System.
UI00123	MRI Scanning Client communication has faulted.	Reboot the console and services computers. If the problem persists, call service.
UI00124	MRI Scanning Client communication has disconnected.	Reboot the console and services computers. If the problem persists, call service.
UI00125	There was an error saving the volume.	The user should check the connection to the database, and retry. If the problem persists, call service.
UI00126	for executable	Call service.
UI00127	There was an error loading the delivered fraction.	The user should check the connection to the database, and retry. If the problem persists, call service.
UI00128	Fraction does not have a delivered plan.	Call service.
UI00129	Could not load plan and/or prescription info.	The user should check the connection to the database, and retry. If the problem persists, call service.
UI00130	Fraction does not exist.	The user should check the connection to the database, and retry. If the problem persists, call service.
UI00131	You must provide a plan name.	Enter a plan name and retry.
UI00132	The plan was not saved. Could not create the new plan.	The user should check the connection to the database, and retry. If the problem persists, call service.
UI00133	There was an error saving the new plan.	The user should check the connection to the database, and retry. If the problem persists, call service.
UI00134	Could not reload original plan info.	The user should check the connection to the database, and retry. If the problem persists, call service.
UI00135	There is no dose to export.	Nothing.
UI00136	The system was unable to store the dose volume.	The user should check the connection to the database, and retry. If the problem persists, call service.

ID	DIALOG MESSAGE	USER ACTION
UI00137	Could not load delivery calendar.	The user should check the connection to the database, and retry. If the problem persists, call service.
UI00138	Could not load patient.	The user should check the connection to the database, and retry. If the problem persists, call service.
UI00139	Could not load patient or diagnosis.	The user should check the connection to the database, and retry. If the problem persists, call service.
UI00140	Could not load prescription.	The user should check the connection to the database, and retry. If the problem persists, call service.
UI00141	Error loading plans.	The user should check the connection to the database, and retry. If the problem persists, call service.
UI00142	The selected plan is empty.	Nothing.
UI00143	Could not assign to fraction.	The user should check the connection to the database, and retry. If the problem persists, call service.
UI00144	Could not lock fraction.	The user should check the connection to the database, and retry. If the problem persists, call service.
UI00145	Could not unlock fraction.	The user should check the connection to the database, and retry. If the problem persists, call service.
UI00146	Could not unassign fractions.	The user should check the connection to the database, and retry. If the problem persists, call service. A workaround may exist by selecting the new plan to be assigned and using the assign function.
UI00147	Could not consolidate or load fraction calendar.	The user should check the connection to the database, and retry. If the problem persists, call service.
UI00148	The number of fractions assigned plans must match the number of fractions stated in the prescription. Please adjust the prescription or the fraction calendar and try again.	Fix the delivery calendar plan assignments to match the prescription(s), and retry.
UI00149	Can Not Approve	The user should check the connection to the database, and retry. If the problem persists, call service.
UI00150	Not all plans in the delivery calendar are approved for treatment. Approve or replace the plans.	The user may approve or replace the unapproved plans, and retry.

ID	DIALOG MESSAGE	USER ACTION
UI00151	Not all prescriptions in the delivery calendar are approved for treatment. Approve the prescription and corresponding plans to continue.	The user may approve the prescription(s) and corresponding plan(s), and retry.
UI00152	Could not approve delivery calendar.	The user should check the connection to the database, and retry. If the problem persists, call service.
UI00153	Could not un-approve delivery calendar.	The user should check the connection to the database, and retry. If the problem persists, call service.
UI00154	Could not insert into delivery calendar.	The user should check the connection to the database, and retry. If the problem persists, call service.
UI00155	Could not delete completion fraction from delivery calendar.	The user should check the connection to the database, and retry. If the problem persists, call service.
UI00156	There was an error loading the cumulative dose.	The user should check the connection to the database, and retry. If the problem persists, call service.
UI00157	There is no dose to export.	The user should check the connection to the database, and retry. If the problem persists, call service.
UI00158	The system was unable to store the dose volume.	The user should check the connection to the database, and retry. If the problem persists, call service.
UI00159	Could not retrieve patient history.	The user should check the connection to the database, and retry. If the problem persists, call service.
UI00160	Could not load delivery calendar data for diagnosis.	The user should check the connection to the database, and retry. If the problem persists, call service.
UI00161	Error loading patient information.	The user should check the connection to the database, and retry. If the problem persists, call service.
UI00162	There was an error setting active/inactive status.	The user should check the connection to the database, and retry. If the problem persists, call service.
UI00163	There was an error loading the plan.	The user should check the connection to the database, and retry. If the problem persists, call service.
UI00164	The selected plan is empty.	Nothing.
UI00165	Unable to delete diagnosis.	The user should check the connection to the database, and retry. If the problem persists, call service.

ID	DIALOG MESSAGE	USER ACTION
UI00166	Unable to delete prescription.	The user should check the connection to the database, and retry. If the problem persists, call service.
UI00167	Unable to delete plan.	The user should check the connection to the database, and retry. If the problem persists, call service.
UI00168	Unable to delete patient.	The user should check the connection to the database, and retry. If the problem persists, call service.
UI00169	Could not load the plan to copy.	The user should check the connection to the database, and retry. If the problem persists, call service.
UI00170	Could not aquire plan write token for newly-created plan.	The user should check the connection to the database, and retry. If the problem persists, call service.
UI00171	Could not save the copy.	The user should check the connection to the database, and retry. If the problem persists, call service.
UI00172	There was an error copying plan.	Retry. If the problem persists, call service.
UI00173	Unable to load drop downs from database.	The user should check the connection to the database, and retry. If the problem persists, call service.
UI00174	Could not save the prescription	The user should check the connection to the database, and retry. If the problem persists, call service.
UI00175	You must select a prescription.	Select a prescription and try operation again.
UI00176	Error creating plan.	The user should check the connection to the database, and retry. If the problem persists, call service.
UI00177	A single fraction must be selected, and the completion fraction will be inserted after this fraction.	Select a single fraction to insert the completion fraction.
UI00178	Only the last delivered fraction or future fractions may have completion fractions added.	Select a different fraction to add the completion fraction to.
UI00179	Maximum number of completion fractions reached.	Do not need to add any additional completion fractions. To add additional fractions, change prescription and plan for patient.
UI00180	A single completion fraction must be selected. Completion fractions have non-integer fraction numbers.	Select a completion fraction to add to the delivery calendar.
UI00181	Delivered fractions may not be deleted.	Cannot remove a delivered fraction from the delivery calendar.

ID	DIALOG MESSAGE	USER ACTION
UI00182	Could not load objective template list.	Try the operation again. If the problem persists, reboot and try again. If the problem still persists, call service.
UI00183	Could not load prescription.	The user should check the connection to the database, and retry. If the problem persists, call service.
UI00184	You will leave the patient management task and enter the planning workflow. Continue?	Confirm to leaving the patient management workflow.
UI00185	Attempted to delete a plan, but the plan could not be determined.	The user should check the connection to the database, and retry. If the problem persists, call service.
UI00186	A plan with that name already exists. Please enter a different name.	Choose a different plan name and retry.
UI00187	Error checking plan name.	The user should check the connection to the database, and retry. If the problem persists, call service.
UI00188	Could not change the plan QA status.	The user should check the connection to the database, and retry. If the problem persists, call service.
UI00189	Error saving objective template.	The user should check the connection to the database, and retry. If the problem persists, call service.
UI00190	An error has occured retrieving data from the database.	The user should check the connection to the database, and retry. If the problem persists, call service.
UI00191	There was an error loading the plan.	The user should check the connection to the database, and retry. If the problem persists, call service.
UI00192	Tried to set an image set with more than one volume to primary.	Image sets with a single volume can be used as primary. Image sets with multiple volumes may not be used as primary image set.
UI00193	Error preparing screen. An error has occured retrieving data from the database.	The user should check the connection to the database, and retry. If the problem persists, call service.
UI00194	Could not search series.	Try the operation again. If the problem persists, reboot and try again. If the problem still persists, call service.
UI00195	Tried to add a series that is already in the plan.	Select a different series to add to plan.
UI00196	The system encountered an error while updating the plan data after the series load completed.	Try the operation again. If the problem persists, reboot and try again. If the problem still persists, call service.

ID	DIALOG MESSAGE	USER ACTION
UI00197	You cannot remove the primary image set.	Select another series to remove. Cannot remove primary image set from plan.
UI00198	You cannot change the primary image set on this plan.	Create a new plan with a different primary image set if a new primary image set is desired.
UI00199	Error un-approving image registration.	Try the operation again. If the problem persists, reboot and try again. If the problem still persists, call service.
UI00200	Could not edit series. Modality not supported.	Only CT, PET, and MRI series may be editted. Select another series and try again.
UI00201	An error occured while attempting to interpolate the images.	Try the operation again. If the problem persists, reboot and try again. If the problem still persists, call service.
UI00202	There was an error saving the edited volume.	Try the operation again. If the problem persists, reboot and try again. If the problem still persists, call service.
UI00203	Select at least one registration type checkbox to perform automatic registration.	Select rigid, deformable, or both registration type and try again.
UI00204	An error occured while performing image registration.	Try the operation again. If the problem persists, reboot and try again. If the problem still persists, call service.
UI00205	A new volume could not be created.	Try the operation again. If the problem persists, reboot and try again. If the problem still persists, call service.
UI00206	An error occured while locking the image registration.	Try the operation again. If the problem persists, reboot and try again. If the problem still persists, call service.
UI00207	New volume could not be created.	Try the operation again. If the problem persists, reboot and try again. If the problem still persists, call service.
UI00208	An error occured while unlocking the image registration.	Try the operation again. If the problem persists, reboot and try again. If the problem still persists, call service.
UI00209	Error approving image registration.	Try the operation again. If the problem persists, reboot and try again. If the problem still persists, call service.
UI00210	Error un-approving image registration.	Try the operation again. If the problem persists, reboot and try again. If the problem still persists, call service.
UI00211	Could not add the structure.	Try the operation again. If the problem persists, reboot and try again. If the problem still persists, call service.

ID	DIALOG MESSAGE	USER ACTION
UI00212	Could not add the structure.	Try the operation again. If the problem persists, reboot and try again. If the problem still persists, call service.
UI00213	Skin is a required structure.	Do not delete skin contour.
UI00214	An error occured during skin contouring.	Try the operation again. If the problem persists, reboot and try again. If the problem still persists, call service.
UI00215	An error occured during skin contouring threshold calculation.	Try the operation again. If the problem persists, reboot and try again. If the problem still persists, call service.
UI00216	Error saving contour template	Try the operation again. If the problem persists, reboot and try again. If the problem still persists, call service.
UI00217	Structure limit reached. Maximum number of user-added structures is {0}. Remove {1} structure(s) before auto-contouring.	Remove a structure to add an additional structure.
UI00218	Invalid structure mask value is present.	Try the operation again. If the problem persists, reboot and try again. If the problem still persists, call service.
UI00219	An error occured while auto-contouring.	Try the operation again. If the problem persists, reboot and try again. If the problem still persists, call service.
UI00220	Single slice volume, no slice thickness.	Use multi-slice volume to perform operation.
UI00221	An error occured while expanding the contour by a margin.	Try the operation again. If the problem persists, reboot and try again. If the problem still persists, call service.
UI00222	An error occured while using boolean operations to contour.	Try the operation again. If the problem persists, reboot and try again. If the problem still persists, call service.
UI00223	An error occured while attempting to center displayed slices on the structure.	Try the operation again. If the problem persists, reboot and try again. If the problem still persists, call service.
UI00224	Invalid contouring orientation detected.	Only axial contours may be imported into plan.
UI00225	Could not paste structure because the copy slice was not set.	Copy structure before pasting.
UI00226	Could not paste all structures because the copy slice was not set.	Copy the contours and retry paste.
UI00227	Error approving contours. Could not approve.	Try the operation again. If the problem persists, reboot and try again. If the problem still persists, call service.
UI00228	Cannot approve/un-approve contours. The plan is loaded read-only.	Read-only plan cannot be modified.

ID	DIALOG MESSAGE	USER ACTION
UI00229	Error un-approving contours.	Try the operation again. If the problem persists, reboot and try again. If the problem still persists, call service.
UI00230	Primary image resolution is {0:F2}x{1:F2}x{2:F2}, and auto-contouring template resolution is {3:F2}x{4:F2}x{5:F2}. OK to continue?	Acknowledge dialog and press OK to continue. Verify the output contours carefully. Select cancel, to cancel auto-contouring operation.
UI00231	The plan must contain at least one isocenter.	Add an isocenter and try operation again.
UI00232	Moving this point outside the feasible region will delete all beam angle groups associated with this isocenter.	Conform moving the isocenter is desired operation.
UI00233	An error occured while centering the point on the selected structure.	Try the operation again. If the problem persists, reboot and try again. If the problem still persists, call service.
UI00234	An error occured while calculating the feasible region for isocenters.	Try the operation again. If the problem persists, reboot and try again. If the problem still persists, call service.
UI00235	An error occured while auto-fitting MLCs.	Try the operation again. If the problem persists, reboot and try again. If the problem still persists, call service.
UI00236	An error occured while checking for interdigitation of MLCs.	Try the operationa again. If the problem persists, reboot and try again. If the problem still persists, call service.
UI00237	Interdigitation is present for beams at the following angles:	Remove interdigitation and try again.
UI00238	The result is not valid and all interdigitated beams will be cleared.	Try the operation again. If the problem persists, reboot and try again. If the problem still persists, call service.
UI00239	At least one beam set to 'on' must be present to calculate dose.	Set a beam to 'on' and try again.
UI00240	A target must be present to optimize dose.	Define a structure as a target to optimize dose.
UI00241	Fixed conformal beam weighting percentages must add up to 100%. Modify inputs and retry.	Ensure beam weighting percentages add up to 100%.
UI00242	A skin contour must be present to calculate dose.	Add back skin contour and re-try operation.
UI00243	Skin must have volume greater than zero to optimize dose.	Readjust skin contour and try again.
UI00244	One or more structures in this plan have zero volume. Proceed?	Select yes to proceed. Select no to abort operation and ajust contours.
UI00245	You must provide a fused CT volume or set structure densities for dose calculation.	Select electron density and try again.

ID	DIALOG MESSAGE	USER ACTION
UI00246	An error occured while optimizing dose.	Try the operation again. If the problem persists, reboot and try again. If the problem still persists, call service.
UI00247	An error occured while predicting dose using Monte Carlo algorithm.	Try the operation again. If the problem persists, reboot and try again. If the problem still persists, call service.
UI00248	At least one beam set to 'on' must be present to normalize dose.	Set one beam to on and normalize again.
UI00249	A skin contour must be present to normalize dose.	Add skin contour and try again.
UI00250	You must provide a fused CT volume or set structure densities for dose calculation.	Set electron density and try operation again.
UI00251	Invalid input to dose normalization routine.	Try the operation again. If the problem persists, reboot and try again. If the problem still persists, call service.
UI00252	An error occured while normalizing the dose.	Try the operation again. If the problem persists, reboot and try again. If the problem still persists, call service.
UI00253	Number of beam segments changed during normalization due to fluence constraints.	Acknowledge informational message.
UI00254	The returned DVH structure cannot be selected in the plan.	Try again. If the problem persists, reboot the system. If the problem still persists, call service.
UI00255	Invalid DVH returned from algorithm engine.	Try again. If the problem persists, reboot and try again. If the problem still persists, call service.
UI00256	Returned Max Dose is greater than {0} Gy. Please reset some parameters and re-optimize.	Change parameters and re-optimize.
UI00257	No isocenter defined.	Choose isocenter and try operation again.
UI00258	Single slice volume, no slice thickness.	Choose a volume with more than one slice for operation.
UI00259	An error occured during the expand by margin operation.	Try the operation again. If the problem persists, reboot and try again. If the problem still persists, call service.
UI00260	Electron density must be set and saved during dose optimization before performing dose prediction.	Select electron density map before optimizing and predicting dose.
UI00261	At least one beam must be present to perform dose prediction.	Add a beam before attempting to optimize and predict dose.
UI00271	An error occured while predicting dose.	Try the operation again. If the problem persists, reboot and try again. If the problem still persists, call service.

ID	DIALOG MESSAGE	USER ACTION
UI00272	Returned Max Dose is greater than {0} Gy. Please reset some parameters and re-optimize.	Adjust plan parameters and attempt optimization again.
UI00273	The returned DVH structure cannot be selected in the plan.	Try again. If the problem persists, reboot the system. If the problem still persists, call service.
UI00274	Invalid DVH returned from algorithm engine.	Try again. If the problem persists, reboot and try again. If the problem still persists, call service.
UI00275	Could not retrieve plans for patient.	Try the operation again. If the problem persists, reboot and try again. If the problem still persists, call service.
UI00276	There was an error approving the plan.	Try the operation again. If the problem persists, reboot and try again. If the problem still persists, call service.
UI00277	There was an error un-approving the plan.	Try the operation again. If the problem persists, reboot and try again. If the problem still persists, call service.
UI00278	Structure name is a required field.	Enter or select a structure name to add a new structure.
UI00279	Unable to load user drop down from database.	Try the operation again. If the problem persists, reboot and try again. If the problem still persists, call service.
UI00280	Could not get contour template list.	Try the operation again. If the problem persists, reboot and try again. If the problem still persists, call service.
UI00281	Error retrieving structures in contour template.	Try the operation again. If the problem persists, reboot and try again. If the problem still persists, call service.
UI00282	You must select a template.	Select a template to auto contour from.
UI00283	You must select at least one structure.	Select at least one structure to auto-contour.
UI00284	Failed retreiving auto-contouring template.	Try the operation again. If the problem persists, reboot and try again. If the problem still persists, call service.
UI00285	Unable to load user drop down from database.	Try the operation again. If the problem persists, reboot and try again. If the problem still persists, call service.
UI00286	You must select a user for the template.	Select a user to save the template under.
UI00287	You must enter a template name.	Select a template name before saving template.
UI00288	You must enter a sex for the template.	Select a sex before saving template.
UI00289	You must select a region for the template.	Select a region before saving template.
UI00290	Maximum number of beam angle groups = 128. Remove a beam angle group and try again.	Remove a beam angle group or decrease the number of added beams and try again.

ID	DIALOG MESSAGE	USER ACTION
UI00291	X1 must be less than X2. Re-enter new values and try again.	Re-enter the values and try again.
UI00292	Maximum number of beam angle groups = 128. Remove a beam angle group or decrease the number of added beams and try again.	Remove a beam angle group or decrease the number of added beams and try again.
UI00293	Unable to check Beam Template Name.	Try the operation again. If the problem persists, reboot and try again. If the problem still persists, call service.
UI00294	Enter a target structure.	Enter a target structure for the resulting structure to be copied to.
UI00295	Enter a structure to be subtracted from.	Enter a structure to subtract from and try again.
UI00296	Enter a structure to be subtracted.	Enter a structure to subtract and try again.
UI00297	Select 2 or more structures to add together.	Select 2 or more structures to add together and try again.
UI00298	Select 2 or more structures to intersect.	Select 2 or more structures to computer intersection.
UI00299	Enter a valid name.	Enter a different name for contour.
UI00300	There was an error loading the comparison plan.	Try the operation again. If the problem persists, reboot and try again. If the problem still persists, call service.
UI00301	Could not approve plan.	Try the operation again. If the problem persists, reboot and try again. If the problem still persists, call service.
UI00302	Cannot approve/un-approve contours. The plan is loaded read-only.	Cannot change approval status for read only plans.
UI00303	All IMRT beams will be converted to optimized conformal beams. There is no undo available. Continue?	Confirm to change all beam types to optimized conformal.
UI00304	Unable to check structure constraints template label.	Try the operation again. If the problem persists, reboot and try again. If the problem still persists, call service.
UI00305	There was an error saving the structure contraints template.	Try the operation again. If the problem persists, reboot and try again. If the problem still persists, call service.
UI00306	There was an error overwriting the structure contraints template.	Try the operation again. If the problem persists, reboot and try again. If the problem still persists, call service.
UI00307	Could not add the new structure.	Try the operation again. If the problem persists, reboot and try again. If the problem still persists, call service.
UI00308	Could not retrieve patient information.	Try the operation again. If the problem persists, reboot and try again. If the problem still persists, call service.

ID	DIALOG MESSAGE	USER ACTION
UI00309	Could not retrieve patients.	Try the operation again. If the problem persists, reboot and try again. If the problem still persists, call service.
UI00310	Could not retrieve Radiation Oncologist list.	Try the operation again. If the problem persists, reboot and try again. If the problem still persists, call service.
UI00311	Could not retrieve Medical Physicist list.	Try the operation again. If the problem persists, reboot and try again. If the problem still persists, call service.
UI00312	Could not retrieve Dosimetrist list.	Try the operation again. If the problem persists, reboot and try again. If the problem still persists, call service.
UI00313	Could not retrieve retrieve Radiation Oncologist, Medical Physicist, and Dosimetrist lists.	Try the operation again. If the problem persists, reboot and try again. If the problem still persists, call service.
UI00314	The image sets cannot be modified because the plan is approved.	Unapprove plan to add image sets.
UI00315	Cannot approve/un-approve image registration. The plan is read-only.	Cannot modify image registration status, i.e. add images for read only plans.
UI00316	Image series patient name, ID, sex, or birthdate does not match the currently loaded patient.	Confirm before adding image series to patient.
UI00317	Click OK to add the series to the plan.	Confirm to add the series to the plan.
UI00318	You cannot modify the image sets because the plan is approved.	Unapprove the plan to modify image sets.
UI00319	Found a null current loader, or current loader had a null loaded volume while setting preview volume.	Try loading the plan again. If the problem persists, reboot and try again. If the problem still persists, call service.
UI00320	Unable to load user drop down from database.	Try the operation again. If the problem persists, reboot and try again. If the problem still persists, call service.
UI00321	Unable to retrieve user templates from database.	Try the operation again. If the problem persists, reboot and try again. If the problem still persists, call service.
UI00322	Additional structures exist in the template. Would you like to load these additional structures?	Select yes to add structures, no to abort adding additional structures.
UI00323	There was an error loading the structure contraints template.	Try the operation again. If the problem persists, reboot and try again. If the problem still persists, call service.
UI00324	Could not retrieve diagnoses for patient.	Try the operation again. If the problem persists, reboot and try again. If the problem still persists, call service.

ID	DIALOG MESSAGE	USER ACTION
UI00325	Could not retrieve prescriptions for patient.	Try the operation again. If the problem persists, reboot and try again. If the problem still persists, call service.
UI00326	Could not load diagnosis.	Try the operation again. If the problem persists, reboot and try again. If the problem still persists, call service.
UI00327	Could not retrieve prescriptions.	Try the operation again. If the problem persists, reboot and try again. If the problem still persists, call service.
UI00328	Select a prescription to continue.	Select a prescription and then retry operation.
UI00329	Error creating plan.	Try the operation again. If the problem persists, reboot and try again. If the problem still persists, call service.
UI00330	Could not assign prescription site.	Try the operation again. If the problem persists, reboot and try again. If the problem still persists, call service.
UI00331	Could not add diagnosis.	Try the operation again. If the problem persists, reboot and try again. If the problem still persists, call service.
UI00332	Could not save the prescription.	Try the operation again. If the problem persists, reboot and try again. If the problem still persists, call service.
UI00333	Could not add prescription.	Try the operation again. If the problem persists, reboot and try again. If the problem still persists, call service.
UI00334	Could not retrieve prescription.	Try the operation again. If the problem persists, reboot and try again. If the problem still persists, call service.
UI00335	You cannot modify image registration because the plan is approved.	Unapprove the plan to change approval status of image registration.
UI00336	Cannot approve/un-approve image registration. The plan is loaded read-only.	Unapprove the plan to change approval status of image registration.
UI00337	Could not retrieve user list.	Try the operation again. If the problem persists, reboot and try again. If the problem still persists, call service.
UI00338	A structure constraints template with that name already exists. Would you like to overwrite it?	Choose yes to overwrite, no to choose another name.
UI00339	There was an error saving the structure constraints template.	Try the operation again. If the problem persists, reboot and try again. If the problem still persists, call service.
UI00340	Could not retrieve structure names.	Try the operation again. If the problem persists, reboot and try again. If the problem still persists, call service.

ID	DIALOG MESSAGE	USER ACTION
UI00341	Please select or create a structure.	Select or create a structure.
UI00342	Could not get structure information.	Try the operation again. If the problem persists, reboot and try again. If the problem still persists, call service.
UI00343	There was an error retreiving users.	Try the operation again. If the problem persists, reboot and try again. If the problem still persists, call service.
UI00344	Administrator username is required. Enter an administrator username and try again.	Enter an administrator's username and try again.
UI00345	Administrator password is required. Enter a password and try again.	Enter admininstrator password to edit user.
UI00346	Password is required. Enter a password and try again.	Enter a password and try again.
UI00347	Entered password values do not match. Re-enter password values and try again.	Re-enter passwords and try again.
UI00348	Username is required. Enter a username and try again.	Enter the username before saving user.
UI00349	User group is required. Select a group and try again.	Select a user group before saving patient.
UI00350	Entered password values do not match. Re-enter password values and try again.	Re-enter the passwords and try again.
UI00351	Username is required. Enter a username and try again.	Enter a username before trying to save the user.
UI00352	Password is required. Enter a password and try again.	Enter the password and try again.
UI00353	User group is required. Select a group and try again.	Select a user group and try again.
UI00354	Could not authenticate user.	Re-enter the credentials and try again.
UI00355	Authentication Failed	Try entering the credentials and try again. If the problem persists, call service.
UI00356	Could not load diagnoses.	Try the operation again. If the problems persists, reboot and try the operation again. If the problem still persists call service.
UI00357	Could not unlock fraction.	Try the operation again. If the problems persists, reboot and try the operation again. If the problem still persists call service.
UI00358	There was an error loading the plan.	Try the operation again. If the problems persists, reboot and try the operation again. If the problem still persists call service.
UI00359	The plan is not approved for treatment.	Ensure the plan is approved and start the delivery process again.

ID	DIALOG MESSAGE	USER ACTION
UI00360	There was an error preparing the loaded plan.	Try the operation again. If the problems persists, reboot and try the operation again. If the problem still persists call service.
UI00361	Could not store setup notes.	Try the operation again. If the problems persists, reboot and try the operation again. If the problem still persists call service.
UI00362	Interdigitation is present for one or more MLC shapes. This plan is undeliverable.	Go to planning and check plan. Call service.
UI00363	Could not save imaging protocol information to database.	Try the scanning operation again. If the problems persists, reboot and try the scanning operation again. If the problem still persists call service.
UI00364	MRI Scanning Error: {0}	Try the scan again. If the problems persists, reboot and try the scan again. If the problem still persists call service.
UI00365	Timeout waiting for MRI acquisition to complete.	Reboot and try the scan again. If the problem still persists call service.
UI00366	MRI scanner returned empty volumes.	Try the operation again. If the problems persists, reboot and try the operation again. If the problem still persists call service.
UI00367	MRI scanner failed returning volumes.	Try the operation again. If the problems persists, reboot and try the operation again. If the problem still persists call service.
UI00368	MRI scanner console has a popup window. Please read and acknowledge the popup on that system.	Go to the MRI console monitor, check the error, dismiss the error, and check the MRI manual for further instructions regarding the error.
UI00369	MRI scanning client communication has faulted.	Reboot the system and try again. If the problem persists call service.
UI00370	There was an error during processing of the acquired image.	Retry the operation. If the problem persists, retake another image and try again. If the problem still persists, call service.
UI00371	You must have contours in the plan to auto-contour here.	Ensure contours are present in original plan and try again.
UI00372	There was an error during auto-contouring of the acquired image.	Retry the operation. If the problem persists, retake another image and try again. If the problem still persists, call service.
UI00373	There was an error during the deformation of the original plan images.	Retry the operation. If the problem still persists, call service.
UI00374	There was an error during the feasible region for isocenters calculation.	Retry the operation. If the problem persists, retake another image and try again. If the problem still persists, call service.
UI00375	There was an error getting the shift from the deformation results.	Retry the operation. If the problem persists, retake another image and try again. If the problem still persists, call service.

ID	DIALOG MESSAGE	USER ACTION
UI00376	There as an error centering the isocenter on the selected structure.	Try again. If the problem persists, try manually moving the isocenter. Report the problem to service.
UI00377	A hard collision has been detected for the following isocenters: {0}. Adjust patient position on couch, or adjust suggested shift, and retry. Hard collisions may not be overridden.	Adjust patient position on couch, or adjust suggested shift, and retry.
UI00378	Collision detected for the following isocenters: {0}. Adjust patient position on couch, or adjust suggested shift, and retry.	Adjust patient position on couch, or adjust suggested shift, and retry.
UI00379	I understand that a collision has been detected by the system, and I am overriding the system to move the couch.	Authenticate to move couch anyway despite possible patient collision with bore.
UI00380	An error occured during the data shifting calculations to account for couch shifting. Acquire a new image.	Acquire a new image and try again.
UI00381	Primary volume has no modality.	Call service.
UI00382	An error occured while predicting dose.	Try again. If the problem persists, reboot the system. If the problem still persists, call service.
UI00383	Returned Max Dose is greater than {0} Gy. This plan may be invalid. Please reset some parameters and re-optimize.	Reset parameters in planning or normalize plan.
UI00384	The returned DVH structure cannot be selected in the plan.	Try again. If the problem persists, reboot the system. If the problem still persists, call service.
UI00385	Invalid DVH returned from algorithm engine.	Try again. If the problem persists, reboot and try again. If the problem still persists, call service.
UI00386	At least one beam set to 'on' must be present to calculate dose.	Turn beams on in planning workflow and reoptimize.
UI00387	A target must be present to optimize dose.	Ensure target exists in plan. If not define target in planning workflow and reoptimize.
UI00388	A skin contour must be present to calculate dose.	Ensure skin contour is present.
UI00389	There was an error while re-optimizing dose.	Try again. If the problem persists, reboot the system. If the problem still persists, call service.
UI00390	Invalid input to dose normalization routine.	Try again. If the problem persists, reboot the system. If the problem still persists, call service.
UI00391	An error occured while normalizing dose.	Try again. If the problem persists, reboot the system. If the problem still persists, call service.
UI00392	Number of beam segments changed during normalization due to fluence constraints.	Acknowledge warning.

ID	DIALOG MESSAGE	USER ACTION
UI00393	Could not store gating settings imaging setup information	Try adjusting gating settings. If the problem persists, reboot the system. If the problem still persists, call service.
UI00394	Single slice volume, no slice thickness.	Reaquire images. If the problem persists, reboot the system. If the problem still persists, call service.
UI00395	An error occured while expanding the contour by a margin.	Try again. If the problem persists, call service.
UI00396	The system failed updating the user interface display from the machine control values.	Reboot the system and try again. If the problem persists, call service.
UI00397	Fluence to beam on time conversion failed.	Call service.
UI00398	The treatment delivery video could not be saved.	Call service.
UI00399	No image was taken on the positioning screen. Fraction record plan will use original plan images and contours.	Dismiss dialog.
UI00400	Unable to begin treatment. Could not save treatment plan data.	Restart the delivery. If the problem persists, reboot the system and try again. If the problem still persists, call service.
UI00401	Unable to begin treatment. Could not create treatment record.	Restart the delivery. If the problem persists, reboot the system and try again. If the problem still persists, call service.
UI00402	Error during post delivery processing.	Call service.
UI00403	Get Treatment Recorded Data	None.
UI00404	Could not get treatment delivery recorded data.	Reboot the system and the system will try to retrieve the data again. If the problem persists, call service.
UI00405	Recorded treatment delivery data ID mismatch.	Call service.
UI00406	Treatment record ID mismatch. Expected ID {0} but got {1}.	Call service.
UI00407	There was an error storing the treatment data.	Reboot the system and the system will try to retrieve the data again. If the problem persists, call service.
UI00408	Could not add the treatment delivery result.	Reboot the system and the system will try to retrieve the data again. If the problem persists, call service.
UI00409	Could not add the treatment delivery result while saving the treatment playback recorded data.	Reboot the system and the system will try to retrieve the data again. If the problem persists, call service.

ID	DIALOG MESSAGE	USER ACTION
UI00410	Could not compute delivered fraction results.	Reboot the system and the system will try to retrieve the data again. If the problem persists, call service.
UI00411	Could not get delivered fraction results.	Reboot the system and the system will try to retrieve the data again. If the problem persists, call service.
UI00412	There was an error calculating the delivered dose.	Reboot the system and the system will try to retrieve the data again. If the problem persists, call service.
UI00413	The returned DVH structure cannot be selected in the plan.	Reboot the system. If the problem still persists, call service.
UI00414	Invalid DVH returned from algorithm engine.	Try again. If the problem persists, reboot and try again. If the problem still persists, call service.
UI00415	There was an error saving the treatment record dose plan.	Call service.
UI00416	Could not add the treatment delivery result for completed fraction.	Call service.
UI00417	There was an error loading the previous cumulative dose.	Reboot the system and the system will try to retrieve the data again. If the problem persists, call service.
UI00418	There was an error processing the cumulative dose.	Reboot the system and the system will try to retrieve the data again. If the problem persists, call service.
UI00419	There was an error while registering the cumulative dose.	Reboot the system and the system will try to retrieve the data again. If the problem persists, call service.
UI00420	There was an error saving the new cumulative dose.	Reboot the system and the system will try to retrieve the data again. If the problem persists, call service.
UI00421	There was an error processing the cumulative dose.	Reboot the system and the system will try to retrieve the data again. If the problem persists, call service.
UI00422	Could not add the treatment delivery result for partial fraction.	Reboot the system and the system will try to retrieve the data again. If the problem persists, call service.
UI00423	Unable to record into patient history.	Record the information in the patient chart. Inform service.
UI00424	Could not compute remaining dose for the completion plan.	Call service.
UI00425	Could not create completion fraction from partial delivery results.	Call service. Do not continue treatment until problem is discussed with service.
UI00426	An error occured while calculating the dose for the completion fraction.	Call service. Do not continue treatment until problem is discussed with service.

ID	DIALOG MESSAGE	USER ACTION
UI00427	The returned DVH structure cannot be selected in the plan.	Reboot the system. If the problem still persists, call service.
UI00428	Invalid DVH returned from algorithm engine.	Call service.
UI00429	A partial fraction was delivered, resulting in generation of a completion plan. Would you like to deliver the completion plan now?	Select "Yes" to deliver the fraction right away, and the system will insert the completion fraction into the delivery calendar. Select No to unapprove the delivery calendar.
UI00430	Maximum number of completion fractions reached for this fraction. You must deliver this fraction later.	Go to delivery calendar to schedule completion fraction.
UI00431	Could not insert a completion fraction into delivery calendar.	Inform service. Go to delivery calendar to schedule completion fraction.
UI00432	Unable to record into patient history.	Add note of delivery to patient. Call service.
UI00433	There was an error saving the completion plan.	Call service.
UI00434	No dose was delivered. Fraction record will not be marked as delivered.	Restart delivery after checking gating settings. Call service if the problem persists.
UI00435	An attempt to start treatment for fraction {0} under diagnosis {1} was unsuccessful. No radiation was delivered.	Reboot the system and try again. If the problem persists, call service.
UI00436	Beam on time of {0} seconds exceeds maximum value of {1} seconds.	Call service.
UI00438	The system cannot connect to the couch.	Reboot system. If the problem persists, call service.
UI00439	Unknown beam delivery status sent for display.	Abort treatment. Deliver completion fraction. If problem persists, call service.
UI00440	The console received invalid beam data and cannot update the display.	Abort treatment. Deliver completion fraction. If problem persists, call service.
UI00441	All interlocks must be cleared to enable treatment. Please clear all interlocks and try again.	Clear all interlocks and e-stops and try again.
UI00442	This plan has a QA status of failed. To continue loading, override and authenticate.	Check with Physics/QA whether or not to delivery plan. Authenticate to delivery plan despite plan marked as QA failed.
UI00443	Authenticate to load a plan that failed QA for patient treatment.	Check with Physics/QA for plan QA status. Enter credentials to deliver plan.
UI00444	This plan has not had QA performed. To continue loading, click OK.	Click OK to continue with delivery process. Click cancel to abort loading for delivery.
UI00445	Could not write to patient history.	Inform service.
UI00446	Images do not show true patient location because couch shift has not been applied.	Shift the couch to match registration results. If couch shift is not desired, ignore warning.

ID	DIALOG MESSAGE	USER ACTION
UI00447	The system could not auto-assign future fractions to the selected plan. The delivery calendar will be unapproved for future fractions.	After delivery, must re-approve fraction calendar after selecting which plan to use for subsequent fractions.
UI00448	The system could not retrieve the original plan information. Check the database connection and try again. If this problem persists, contact Service.	Reboot the system and try again. If the problem persists, call service.
UI00449	The original plan is selected for viewing, but the re-optimized new plan has been approved for delivery. Continue with re-optimized plan?	Select yes to continue with re-optimized plan. Select no to abort moving to next screen and change plan.
UI00450	The new re-optimized plan is selected for viewing, but has not been approved for delivery. Continue with original plan?	Select yes to continue with original plan. Select no to abort moving to next screen and change plan.
UI00451	The dose has been re-optimized to a new plan, but has not been approved for delivery. Continue with original plan?	Approve new plan if re-optimized plan is desired for treatment, otherwise continue.
UI00452	The predicted dose does not meet prescription constraints. Continue?	Select yes to continue with original plan. Select no to adjust patient position and predict dose again or re-optimize plan.
UI00453	Authenticate to approve the new re-optimized plan for delivery. The new plan will be approved and used for this fraction, and it will be assigned to all future fractions in the delivery calendar which have the same original plan.	Type in credentials to authenticate and use new plan.
UI00454	Authenticate to revert to the original plan for this and all applicable future fractions.	Authenticate to use original plan for this and all subsequent fractions.
UI00455	System could not save the new plan. Check your database connection and try again. If this issue persists, call service.	Exit out of delivery. Restart delivery and try again. If the problem persists, reboot and try again. If the problem still persists call service.
UI00456	There was an error saving the plan.	Exit out of delivery. Restart delivery and try again. If the problem persists, reboot and try again. If the problem still persists call service.
UI00457	Constraint structure not defined	Fix the prescription constraint so that the structure is defined.
UI00458	Constraint structure display paramters not defined	Select different display parameters when adding the structure to the prescription
UI00459	Min > Max	Fix the prescription constraint so that the min is less than the max
UI00460	Mean > Max	Fix the prescription constraint so that the mean is less than the max.
UI00461	Mean < Min	Fix the prescription constraint so that the mean is greater than the min.

ID	DIALOG MESSAGE	USER ACTION
UI00462	Max < Primary Rx Dose for same structure	Fix the prescription constraint so that the max is greater than the primary prescription dose.
UI00463	Min > Primary Rx Dose for same structure	Fix the prescription constraint so that the min is less than the primary prescription dose.
UI00464	comparison operator is missing from the dose to volume constraint	Fix the prescription constraint so that it has a comparison operator for dose to volume.
UI00465	comparison operator is missing from the mean constraint	Fix the prescription constraint so that it has a comparison operator for the mean.
UI00466	mean value is missing from the mean constraint	Fix the prescription constraint so that it has a dose value for the mean.
UI00467	percent volume is missing from the dose to volume constraint	Fix the prescription constraint so that it has a percent volume entry for dose to volume.
UI00468	dose is missing from the dose to volume constraint	Fix the prescription constraint so that it has a dose entry for dose to volume.
UI00469	dose to volume Dose value > Max dose	Fix the prescription constraint.
UI00470	Constraint point name not defined	Fix the prescription constraint.
UI00471	Min > Max	Fix the prescription constraint.
UI00472	Max < Primary Rx Dose for same point	Fix the prescription constraint.
UI00473	Min > Primary Rx Dose for same point	Fix the prescription constraint.
UI00474	Min > Max for duplicate constraints for the same structure/point	Fix the prescription constraints.
UI00475	Min > Max for duplicate constraints for the same structure/point	Fix the prescription constraint.
UI00476	Dose to Volume > Max Dose for duplicate constraints for the same structure/point	Fix the prescription constraint.
UI00477	Dose to Volume > Max Dose for duplicate constraints for the same structure/point	Fix the prescription constraint.
UI00478	Dose to Volume < Min Dose for duplicate constraints for the same structure/point	Fix the prescription constraint.
UI00479	Dose to Volume < Min Dose for duplicate constraints for the same structure/point	Fix the prescription constraint.
UI00480	Data Entry Error	Check all the parameters an try again.
UI00481	Check that all constraints are complete and consistent.	Check constraints to ensure they are complete and do not conflict.
UI00482	Initialization of Data Model failed during initialization. The system is shutting down.	Reboot the system. If the problem persists, call service.
UI00482	This application has not been activated and will be shut down. Call Service.	Call service.
UI00483	Verification of Visage Visualization license failed during initialization. The System is shutting down. Please check the USB Hasp dongle is securely plugged into a USB port.	Check that the dongle is securely plugged into the rear USB port. Reboot. If the problem persists, call service.

ID	DIALOG MESSAGE	USER ACTION
UI00484	The system will now shut down a hanging instance of the application.	The software incorrectly shutdown previously. Report the issue to service.
UI00485	A default structure with that name already exists. Please rename and try again.	Select a different name for the structure and try again.
UI00486	Default Structure Name in Use	Select a different name. If the problem persists, call service.
UI00487	Unable to save default structure to database	Try saving the structure again. If the problem persists, reboot and try again. If the problem still persists, call service.
UI00488	A default structure with the name, '{0}', already exists. Please rename and try again.	Select a different name for the structure and try again.
UI00489	Unable to save default structure to database	Try saving the structure again. If the problem persists, reboot and try again. If the problem still persists, call service.
UI00490	Are you sure you want to remove the '{0}' default structure?	Choose yes to confirm removing the structure. Choose no to abort the operation.
UI00491	Unable to remove the default structure from the database	Try deleting the structure again. If the problem persists, reboot and try again. If the problem still persists, call service.
UI00492	Unable to save default structure to database	Try saving the structure again. If the problem persists, reboot and try again. If the problem still persists, call service.
UI00493	Could not retrieve default structure records.	Reboot the system and try accessing list again. If the problem persists, call service.
UI00494	The value of {0} for {1} is invalid. Verify that the value is within Min/Max range and the type is correct. The type expected is '{2}'.	Ensure data entered is in correct range.
UI00495	Parameter error in calibration data.	Check parameter and try again. If the problem persists call service.
UI00496	Authenticate to enter Service Mode. Once authenticated changes to calibration data are permitted.	Only service users can make changes to calibration data.
UI00497	Error saving parameter changes.	Try saving the parameter again. If the problem persists, reboot and try again. If the problem still persists, call service.
UI00498	Error saving to PDF.	Try saving the file again. If the problem persists, reboot and try again. If the problem still persists, call service.
UI00499	Error saving to text file.	Try saving the file again. If the problem persists, reboot and try again. If the problem still persists, call service.

ID	DIALOG MESSAGE	USER ACTION
UI00500	The file, '{0}', could not be found.	Call service.
UI00501	Error saving parameter data to text file, '{0}'. {1}	Try saving the file again. If the problem persists, reboot and try again. If the problem still persists, call service.
UI00502	Error saving parameter data to PDF file, '{0}'. {1}	Try saving the file again. If the problem persists, reboot and try again. If the problem still persists, call service.
UI00503	An error occured opening the isodose template window.	Reboot the system and try loading template again. If the problem persists, call service.
UI00504	Unable to retrieve template for this user.	Reboot the system and try retrieving template again. If the problem persists, call service.
UI00505	An error occured getting isodose templates by user.	Reboot the system and try retrieving template again. If the problem persists, call service.
UI00506	An error occured loading isodose template data.	Reboot the system and try retrieving template again. If the problem persists, call service.
UI00507	An isodose template with that name already exists. Would you like to overwrite it?	To overwrite select yes, to enter a new name, select no.
UI00508	Unable to save isodose template data.	Reboot the system and try saving template again. If the problem persists, call service.
UI00509	Unable to check isodose line template name against existing library of names.	Reboot the system and try saving template again. If the problem persists, call service.
UI00510	Unable to delete isodose line template.	Reboot the system and try deleting template again. If the problem persists, call service.
UI00511	Unable to check display palette name against existing library of names.	Reboot the system and try saving template again. If the problem persists, call service.
UI00512	An display palette template with that name already exists. Would you like to overwrite it?	Select yes to overwrite, no to use a different name.
UI00513	Unable to delete display palette.	Reboot the system and try deleting template again. If the problem persists, call service.
UI00514	Unable to save display palette template data.	Reboot the system and try saving template again. If the problem persists, call service.
UI00515	Unable to save display palette template data.	Reboot the system and try saving template again. If the problem persists, call service.
UI00516	An error occured while opening structure display palette window.	Reboot the system and try again. If the problem persists, call service.
UI00517	Unable to retrieve display palette for this user.	Reboot the system and try retrieving palette again. If the problem persists, call service.
UI00518	Problem retreiving structure display palettes by user.	Reboot the system and try retrieving palette again. If the problem persists, call service.
UI00519	Unable to load display palette data.	Reboot the system and try retrieving palette again. If the problem persists, call service.

ID	DIALOG MESSAGE	USER ACTION
UI00520	Error loading display palette.	Reboot the system and try retrieving palette again. If the problem persists, call service.
UI00521	Unable to retrieve beam templates.	Reboot the system and try retrieving template again. If the problem persists, call service.
UI00522	System could not close the tracking points file stream writer. Tracking point information may be incomplete for this isocenter.	Reboot the system and report the problem to service.
UI00523	System could not open the file to write tracking point values. Tracking point information will not be available for this isocenter.	Report the problem to service.
UI00524	System could not write tracking point values. Tracking point information may be incomplete for this isocenter.	Report the problem to service.
UI00525	System could not save tracking point record files to the database for isocenter number {0}. Tracking point information may not be available for this isocenter.	Report the problem to service.
UI00526	System could not load tracking point record files from the database.	Reboot the system and try accessing again. If the problem persists, call service.
UI00527	Error saving template.	Reboot the system and try saving again. If the problem persists, call service.
UI00528	Problem getting structure constraint template by user.	Reboot the system and try accessing again. If the problem persists, call service.
UI00529	Unable to delete structure constraint template.	Reboot the system and try deleting again. If the problem persists, call service.
UI00530	Unable to check structure constraint template name against the library of existing names.	Reboot the system and try saving again. If the problem persists, call service.
UI00531	Problem saving structure constraint template.	Reboot the system and try saving again. If the problem persists, call service.
UI00532	Unable to delete contour template from database.	Reboot the system and try deleting again. If the problem persists, call service.
UI00533	Unable to check contour template name against the library of existing names.	Reboot the system and try saving again. If the problem persists, call service.
UI00534	Problem saving contour template.	Reboot the system and try saving again. If the problem persists, call service.
UI00535	Problem getting objective template by user.	Reboot the system and try retrieving template again. If the problem persists, call service.
UI00536	Unable to delete objective template from database.	Reboot the system and try deleting again. If the problem persists, call service.
UI00537	Unable to check objective template name against the library of existing names.	Reboot the system and try saving again. If the problem persists, call service.

ID	DIALOG MESSAGE	USER ACTION
UI00538	Problem saving objective template.	Reboot the system and try saving again. If the problem persists, call service.
UI00539	Unable to save template data to database.	Reboot the system and try saving again. If the problem persists, call service.
UI00540	Expected 1 or 3 treatment cines, but found {0}. If more than 3, only the first 3 will be displayed.	Report the problem to service.
UI00541	The system encountered an error while saving the treatment videos.	Report the error to service.
UI00542	File for {0} not found. Video will not be saved.	Report the error to service.
UI00543	This application configured to use {2} Database version={0}, but the version of the connected Database is={1}. Application will be shut down. Call your service group.	Call service.
UI00544	Could not retrieve race/ethnicity list.	Reboot the system. Check database connection. Try the operation again. If the problem persists, call service.
UI00545	System could not end treatment delivery session. System will shut down.	Reboot the system. If the problem persists, call service.
UI00546	System could not retrieve and store the recorded delivery data from the machine controller. System will shut down.	Reboot the system. The system shall retry retrieving and storing the data. If the problem persists, call service.
UI00547	System timed out while processing post-delivery data with {0} minute time limit. System will shut down.	Reboot the system. The system shall retry processing the data. If the problem persists, call service.
UI00548	Unable to retrieve MRI QA Data from database.	Reboot the system and try again. If the problem persists, call service.
UI00549	Problem encountered loading SNR/ Uniformity screen.	Reboot the system and try again. If the problem persists, call service.
UI00550	Problem encountered saving MRI QA Data.	Reboot the system and try again. If the problem persists, call service.
UI00551	System could not retreive unprocessed delivery data from the database. Some delivery records may be incomplete.	Call service.
UI00552	System could not acknowledge retrieval of delivery information from RTC. Call service.	Call service.
UI00553	The system has detected {0} incomplete treatment delivery records and must connect to RTC to retrieve treatment data. Processing of delivery data will start in {1} seconds.	Wait for processing of delivery data to be completed before using the system.

ID	DIALOG MESSAGE	USER ACTION
UI00554	The system has detected {0} incomplete treatment delivery records. Processing of delivery data will start in {1} seconds.	Wait for processing of delivery data to be completed before using the system.
UI00555	Fluence in Gy/min for Head {0} has not been entered in Physics mode. Patient treatment is not allowed until the appropriate QA has been completed and entered into the system.	Contact service before attempting to use the system for treatment delivery.
UI00556	Fluence in Gy/min for Head {0} has not been entered in Physics mode. The system will use estimated beam on times based on the certificate entry values.	Contact service before attempting to use the system for treatment delivery.
UI00557	User '{0}' does not have sufficient permissions for this operation.	Try again using correct credentials.
UI00558	Ram size check failed for {0} computer. Application will not run. Contact service.	Call service.
UI00559	The system encountered an error while retrieving the gating latency treatment list. Please try again later.	Leave the screen and try again.
UI00597	Could not load treatment video	Report the error to service.
UI00560	The system encountered an error while retrieving the gating latency data. Please try again later.	Export the record again.
UI00561	The system could not read the gating latency data.	Try again. If the problem presists, re-run the test.
UI00562	Only users with sufficient permissions may enter Service Mode. Please enter credentials to check authorization.	Try again with the correct username and password.
UI00563	Only authorized service users with sufficient permissions may launch the Log Viewer. Please enter credentials to check authorization.	Try again with the correct username and password.
UI00564	The data could not be stored.	Try again. If the problem presists, re-run the test.
UI00565	Check for Database version was not successful. Message: {0}. Application will be shut down. Call your service group.	Call service.
UI00566	The system could not run the MLC speed test.	Reboot the system and execute the test again. If the problem persists, call service.
UI00567	The gating latency test failed to retreive the data.	Reboot the system and execute the test again. If the problem persists, call service.
UI00568	The gating latency test failed to save to the database.	Reboot the system and execute the test again. If the problem persists, call service.
UI00569	The gating latency test failed to process.	Reboot the system and execute the test again. If the problem persists, call service.

ID	DIALOG MESSAGE	USER ACTION
UI00570	The primary prescription does not have a dose value.	Set a non-zero dose value for the primary prescription.
UI00571	The primary prescription does not have a primary reference.	Set a primary prescription reference: a point or a structure.
UI00572	The primary prescription type has not been selected.	Set a primary prescription type: point or structure.
UI00573	The primary prescription does not have a percent value.	Set a percentage value for the primary prescription.
UI00574	The number of fractions is set to zero.	Set the number of fractions to a number greater than 0.
UI00575	The prescription cannot be approved for the following reasons:	For each reason in the dialog, enter the required prescription values.
UI00576	The number of fractions in this prescription is set to 1. This means the total plan dose will be delivered in a single fraction. Continue with approval?	Ensure number of fractions is set correctly.
UI00577	Anatomy gating cannot be performed if imaging during treatment is not allowed.	Change gating or imaging setting.
UI00578	Anatomy gating cannot be performed if setup imaging is not allowed.	Change imaging settings.
UI00579	Imaging during treatment cannot be performed if setup imaging is not allowed.	Change imaging settings.
UI00580	Dose prediction cannot be performed if setup imaging is not allowed.	Change prediction setting or setup imaging selection.
UI00581	Dose reoptimization cannot be performed if setup imaging is not allowed.	Change optimization setting or setup imaging selection.
UI00582	One or more gating structures was not fully contained by their bounding structures. This may cause treatment to gate indefinitely for that isocenter. You can either continue with treatment as is, or go back to the gating screen and correct the contours. Would you like to continue treatment as is?	Either select Yes to continue treatment with the current gating contours, or click No and correct the contours.
RT00001	RTC reported fault: a fault was caused by physical hardware safety loop. Either a door was opened during therapy, or a non-door interlock was detected.	An error has occurred in the control system. Call service.
RT00002	RTC reported fault: acquisition pulses are not coming at the expected rate.	An error has occurred in the imaging system. Reboot the system. If the problem persists, call service.
RT00003	RTC reported fault: Directory for dose could not be opened. The physical device may need repair.	An error has occurred in the control system. Call service.

ID	DIALOG MESSAGE	USER ACTION
RT00004	RTC reported fault: Directory for dose could not be traversed by the system. The physical device may need repair.	An error has occurred in the control system. Call service.
RT00005	RTC reported fault: Directory for dose could not be closed. The physical device may need repair.	An error has occurred in the control system. Call service.
RT00006	RTC reported fault: Directory for dose could not be created. The physical device may need repair.	An error has occurred in the control system. Call service.
RT00007	RTC reported fault: Directory for dose could not be opened to build the treatment delivery data file. The physical device may need repair.	An error has occurred in the control system. Call service.
RT00008	RTC reported fault: Directory for treatment delivery data could not be traversed due to file system error. The physical device may need repair.	An error has occurred in the control system. Call service.
RT00009	RTC reported fault: could not open, seek, reed, close an individual treatment delivery data record file. The physical device may need repair.	An error has occurred in the control system. Call service.
RT00010	RTC reported fault: after sending treatment delivery data to console, RTC could not delete the directory and contents for the patient.	An error has occurred in the control system. Call service.
RT00011	RTC reported fault: Dose information is on RTC disk from previous patient.	An error has occurred in the control system. Call service.
RT00012	RTC reported fault: While creating treatment delivery data record to send to the console, the system detected that the data size will exceed the maximum size.	An error has occurred in the control system. Call service.
RT00013	RTC reported fault: Unplanned disconnect detected between RTC and operator's console. The disconnect may or may not have ended a treatment.	An error has occurred in the control system. Call service.
RT00014	RTC reported fault: console sent a request for a planned disconnect to RTC while in the middle of a treatment delivery.	An error has occurred in the control system. Call service.
RT00015	The control system has faulted. Call a service engineer.	An error has occurred in the control system. Call service.
RT00016	RTC reported fault: A packet (in base GPP layers) failed CRC check. The command number may have been wrong and treatment will be terminated.	An error has occurred in the control system. Call service.

ID	DIALOG MESSAGE	USER ACTION
RT00017	RTC reported fault: A plan was downloaded while the table was moving.	An error has occurred in the control system. Call service.
RT00018	RTC reported fault: A plan was downloaded, but the table is not at isocenter.	An error has occurred in the control system. Call service.
RT00019	RTC reported fault: Unknown message. Call service.	An error has occurred in the control system. Call service.
RT00020	RTC reported fault: There is a power failure on the MLCs.	An error has occurred in the control system. Call service.
RT00021	RTC reported fault: The sensors for MLC position have faulted.	An error has occurred in the control system. Call service.
RT00022	RTC reported fault: The MLC system is overloaded.	An error has occurred in the control system. Call service.
RT00023	RTC reported fault: The SMC system has experienced a power failure.	An error has occurred in the control system. Call service.
RT00024	RTC reported fault: Slow movement has been detected on the SMC system.	An error has occurred in the control system. Call service.
RT00025	RTC reported fault: The maximum gating frequency threshold has been exceeded.	An error has occurred in the control system. Call service.
RT00026	RTC reported fault: The hardware watchdog timer has expired or faulted.	An error has occurred in the control system. Call service.
RT00027	RTC reported fault: The couch position has reached a hard system limit.	An error has occurred in the patient table control system. Call service.
RT00028	RTC reported fault: The gantry position has reached a hard system limit.	An error has occurred in the control system. Call service.
RT00029	RTC reported fault: The power supply has faulted.	An error has occurred in the control system. Call service.
RT00030	RTC reported fault: A command was received with a bad checksum.	An error has occurred in the control system. Call service.
RT00031	RTC reported fault: An interlock has been opened during treatment delivery.	Reset e-stop and any manual interlocks. Reboot system. If problem persists or it is unknown why the interlock occurred, call service.
RT00032	A source is stuck in the ON position.\n\nRemove the patient from the system immediately.\n\nThis fault is only clearable by ViewRay Service.	A critical error has occurred. Remove the patient and close the vault. Call service immediately.
RT00033	\n\nThis fault is only clearable by ViewRay Service.	A critical error has occurred. Remove the patient and close the vault. Call service immediately.

INDEX

INDEX

VIEWRAY

ViewRay System Treatment Planning and Delivery System

The ViewRay™ Treatment Planning and Delivery System (TPDS) software enables clinicians using the ViewRay system to confidently develop radiotherapy treatment plans and deliver treatment to the target as prescribed, while minimizing dose to surrounding healthy tissue. Advanced planning and delivery tools speed contouring, dose prediction, and dose optimization for streamlined plan creation and evidence-based plan adaptation.

Using a unique combination of magnetic resonance imaging (MRI) and radiotherapy technology, the ViewRay system provides continuous soft-tissue imaging during the delivery of radiotherapy treatment. The system is designed so that clinicians will be able to see where the actual radiation dose is being delivered and adapt to changes in the patient's anatomy. ViewRay system software effectively coordinates imaging, patient positioning, motion control, and treatment delivery.

Highlights

- ***Advanced treatment delivery.*** Support for advanced and conventional radiotherapy techniques, including image-guided radiotherapy (IGRT), stereotactic radiosurgery (SRS) and stereotactic radiotherapy (SRT), intensity modulated radiation therapy (IMRT), and 3D conformal therapy.
- ***Ultrafast autocontouring.*** Sophisticated contouring tools to segment organs-at-risk and accurately delineate target volumes in minutes.
- ***On-table dose prediction.*** Just-in-time dose prediction, using volumetric image data captured just prior to treatment to detect organ motion.
- ***Progressive planning.*** Dose reoptimization for improved accuracy over the course of treatment. This evolutionary planning capability includes on-table reoptimization based on dose predictions, and offline plan review and reoptimization based on the recording of the actual delivered dose.
- ***Real-time 4D motion control.*** Continuous tracking of soft tissue and automatic beam gating for safe and accurate motion control during treatment.
- ***Stringent quality assurance.*** Built-in QA features to help ensure the safety and accuracy of treatment delivery.
- ***User-friendly interface.*** An intuitive, easy-to-use graphical user interface that follows natural clinical workflows.

Support for advanced treatment techniques

The powerful ViewRay system and its software together support advanced radiotherapy techniques such as image-guided radiotherapy (IGRT) and stereotactic radiotherapy and radiosurgery (SRT and SRS), as well as conventional techniques such as 3D conformal

and intensity-modulated radiation therapy (IMRT). With the ViewRay system, clinicians can plan and deliver precision radiotherapy and radiosurgery to treat lesions, tumors, and conditions anywhere in the body where radiation treatment is indicated.

The MRI-guided ViewRay system is the first radiotherapy delivery system capable of providing a stream of high-quality MR patient images during beam-on irradiation. Unlike volumetric CT and X-ray tools from competitors, the ViewRay system's real-time patient imaging provides better soft-tissue contrast without exposing the patient to unnecessary ionizing radiation. This imaging capability, combined with advanced motion control systems and dose prediction and reoptimization capabilities, gives clinicians greater control over the patient's therapy. Because the ViewRay system makes it possible to see and respond to changes in soft tissue before, during, and after a treatment is delivered, clinicians can make more informed decisions and create more finely tuned treatment plans.

Fast, accurate treatment planning and delivery

The ViewRay™ Treatment Planning and Delivery System (TPDS) software allows clinicians to create and deliver treatment plans and to perform quality assurance (QA) procedures following the familiar steps of the treatment planning process: image acquisition and input, anatomy definition, dose calculation, and plan evaluation, implementation, and review.

Because the ViewRay system captures real-time volumetric images and the TPDS software allows ultrafast auto-contouring, dose prediction, and dose optimization, clinicians can verify and adjust plans while the patient is on the treatment table, ensuring that the prescribed dose is aimed at the position of the target. And because the ViewRay system records the actual delivered dose and compiles the cumulative dose delivered over the course of therapy, clinicians can make evidence-based decisions when adapting treatment plans between fractions.

Ultrafast auto-contouring

Sophisticated contouring tools enable users to segment organs-at-risk and accurately delineate target volumes in just minutes. Users define targets and critical structures on fused multimodality images, using advanced drawing and editing capabilities.

Both manual and automatic contouring techniques can be used to define targets and critical structures. Ultrafast auto-contouring algorithms use either standard provided templates or custom templates. The system can also use an existing image and corresponding structure set for each patient to allow for daily auto-contouring of newly acquired image data.

On-table dose prediction

Just-in-time dose prediction, using volumetric image data captured just prior to treatment to detect organ deformation and motion, helps ensure plan accuracy. While the patient is on the treatment table, the ViewRay system can quickly capture a high-quality soft-tissue image, compare it to the planning image, and predict the dose to be

delivered. If needed, clinicians can adapt plans on-the-fly to improve daily dose conformality.

On-table dose prediction provides additional quality assurance and gives clinicians the accurate information they need to make treatment decisions. Because images are captured with an MRI system, the process can be completed safely on a daily basis with no ionizing radiation delivered to the patient.

Progressive planning

With the ViewRay TPDS software, clinicians can easily reoptimize dose and continually adapt plans for improved accuracy over the course of therapy as the tumor changes shape and the patient's internal organs shift. This evolutionary capability enables both online and offline plan adaptation.

Online adaptation. While the patient is on the treatment table, ultrafast dose calculations allow plans to be quickly adapted and reoptimized based on pretreatment dose predictions that account for organ deformation and motion.

Offline adaptation. During treatment, ViewRay software tracks the target and organs-at-risk and records the delivery parameters. Physicians can later review the delivered dose achieved by combining the pre-treatment volumetric image with the delivery information, along with the cumulative dose delivered, to inform offline plan adaptation between fractions. This data helps ensure that dose is delivered as prescribed and allows the prescription to be adjusted, if needed, based on evidence of the actual dose delivered to the target and critical structures.

Real-time 4D motion control

Sophisticated methods for tracking organ motion and controlling the treatment beam promote accurate treatment delivery while further protecting critical structures.

With the ViewRay system, safe and accurate motion control is achieved through real-time tracking of soft tissue and automatic beam gating. With this unique MRI-guided system, three orthogonal soft-tissue imaging planes can be captured continuously during treatment without delivering ionizing radiation to the patient. If the tumor or a critical structure moves beyond a user-defined boundary, the treatment beam is automatically paused; when the structure moves back into the target zone, treatment automatically resumes. The TPDS software allows clinicians to set both spatial and time thresholds for pausing treatment delivery.

No other treatment system is able to control for the motion of a patient's organs based on real-time 4D MRI data.

Stringent quality assurance

The ViewRay TPDS software includes multiple QA features to help ensure the safety and accuracy of treatment delivery. For example:

- Every treatment plan can be verified in a plan QA function that is integrated with the TPDS software.

- The software requires that all contours—whether they are manually generated or automatically generated—be reviewed, validated, and approved by an authorized user.
- The physician specifies the patient prescription, including dose constraints, technique, and tolerance levels for coverage volume, hotspots, and cold spots on a target. The physician may also prescribe whether dose prediction, on-table reoptimization, and soft-tissue gating are allowed, mandatory, or optional.
- Because the TPDS software allows for tolerances in the prescription, differences in the planned and actual DVH values that exceed specified thresholds in doses or volumes will cause the system to issue a warning to the clinician.

User-friendly interface

The ViewRay software features an easy-to-use graphical user interface that keeps learning curves short. The software is designed to follow a natural clinical workflow for intuitive navigation and efficient use.

For increased departmental efficiency, additional workstations can be added to allow for contouring, physician review, and physics QA and approval. If workstations are distributed throughout a clinic, plans can be evaluated anywhere and anytime, speeding the planning process with multiple dose displays and electronic plan approval.

Further information

To learn more about the ViewRay™ Treatment Planning and Delivery System or software, contact your local ViewRay representative or visit http://www.viewray.com.

VARIAN medical systems | A partner for life

Oncology | Radiation Oncology
Eclipse | Overview

- ARIA
  Information System
- Eclipse
  Treatment Planning
- Acuity
  Verification & Simulation
- Trilogy
  Treatment Delivery
- Clinac
  Treatment Delivery
- Upgrades

Eclipse | Overview

Eclipse™ is a comprehensive treatment planning system that simplifies modern radiation therapy planning for all kinds of treatment, including 3D conformal, intensity-modulated radiation therapy (IMRT), electron, proton, and brachytherapy.

With the rich functionality in Eclipse, dosimetrists, physicists, and physicians can efficiently create, select, and verify the best treatment plans for their patients. In addition to ensuring high standards of care and effective protocols, Eclipse provides clinicians with the flexibility to quickly tailor plans for each patient.

Designed to meet the needs of modern clinics and evolving technologies, Eclipse supports advanced processes such as image-guided radiation therapy (IGRT) and Dynamic Adaptive Radiation Therapy (DART™). The efficiency and cost effectiveness of Eclipse enables growing clinics to adopt advanced techniques, protecting investments while improving the quality of care.

In February, Varian shipped its 10,000th Eclipse Treatment Planning System. (Click here for the story)

Contact Varian Oncology

Varian Oncology Headquarters
Tel: 1.650.424.5700
Contact Us

Events

- **08/08/2010 - 08/13/2010**
  CAARI 21st International Conference on the Application of Accelerators in Research & Industry
- **08/08/2010 - 08/21/2010**
  15th Brazilian Congress of Medical Physics
- **08/11/2010 - 08/13/2010**
  MOGA - Medical Oncological Group of Australia
- **08/18/2010 - 08/21/2010**
  World Cancer Congress (International Union Against Cancer)

VARIAN medical systems | A partner for life

Oncology | Radiation Oncology
Eclipse | Treatment Techniques

- **ARIA**
  Information System
- **Eclipse**
  Treatment Planning
- Acuity
  Verification & Simulation
- **Trilogy**
  Treatment Delivery
- Clinac
  Treatment Delivery
- Upgrades

Eclipse | Treatment Techniques

Eclipse is a comprehensive solution for treatment planning that allows clinicians to quickly customize treatment plans for any disease site with precision and accuracy.

Designed for flexibility and feature-rich efficiency, Eclipse simplifies advanced treatment techniques, enabling growing clinics to offer improved patient care and services.

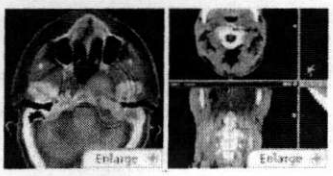
RapidArc™ Radiotherapy Technology

RapidArc™ radiotherapy technology integrates seamlessly with Eclipse to enable clinicians to develop IMRT-quality treatments with a single arc—even for complex treatments.

By combining the power of Eclipse planning tools with the precision and speed of RapidArc, clinicians can develop highly individualized treatments that:

- take one-half to one-eighth the time of similar treatment on other RT platforms - just two minutes in many cases.
- establish an entirely new benchmark for treatment speed and patient comfort
- excel in covering target goals while sparing critical structures

Interactive IMRT Planning



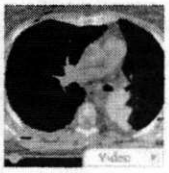
Powerful 3D conformal planning tools in Eclipse are combined with interactive dose-volume optimization for fast, flexible, and accurate intensity-modulated radiation therapy (IMRT) planning. Automatic optimization of beam geometry simplifies the task of selecting the best beam angles for IMRT.

Clinicians can make real-time decisions using the interactive IMRT features in Eclipse. With thousands of patients at hundreds of institutions worldwide successfully treated with IMRT plans generated by Eclipse, it is an established solution for IMRT planning.

- Rapid IMRT Plan Setup
- Clinical Flexibility
- Interactive Optimization
- Accurate Dose Calculation
- Beam Angle Optimization

IGRT

Image-Guided Radiation Therapy (IGRT) allows the user to track changes in patients during the course of treatment. For example, treating a head and neck tumor in the Varian® Inspiration™ environment, the clinician can use integrated Cone-Beam CT (CBCT) planning and fast calculation in Eclipse, to facilitate adaptive therapy (e.g. Dynamic Adaptive Radiation Therapy (DART™)).



4D Planning (Optional)

The Eclipse™ Treatment Planning System's 4D functionality includes contouring and field set up, which is as fast and efficient as in 3D planning, with the added benefit of 4D visualization throughout the planning process. Eclipse displays the motion of targets and critical structures using specially designed 4D tools.

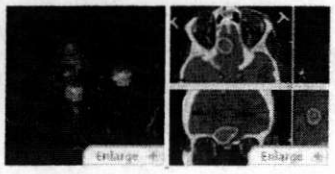
With this visual information, the clinician can easily design 3D conformal and IMRT treatments using images that are either retrospectively binned according to the Real-time Position Management™ (RPM) system signal or prospectively acquired using RPM gated imaging. For retrospective 4D images, Eclipse automatically registers phase-or amplitude-binned image series together with any corresponding derived image series such as MIP, Min-IP, Average-IP or Free Breathing images.

The clinician can view and assess the motion by displaying the 4D image series as movie loops and as blended or "blinking" images. The 4D image display accommodates CT, PET/CT, PET, MR and Cone-Beam CT from the On-Board Imager®. The 4D images are displayed in 2D, 3D, and digitally reconstructed radiograph (DRR) views. Doses for gated treatments are readily visualized in Eclipse.

Adaptive Therapy

Dynamic Adaptive Radiation Therapy (DART) is an exciting, new, emerging capability that has the potential to improve clinical outcomes with relatively little effect on the current level of clinical productivity. In Varian's integrated oncology environment, all parameters required for patient treatment are immediately available to the electronic medical record and treatment delivery systems without importing or exporting. Eclipse can reconstruct dose distributions based upon CBCT images and actual treatment parameters for DART.

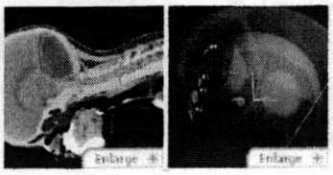
Varian Stereotactic Planning



Stereotactic planning software from Varian offers a complete treatment planning solution for all accelerator-based stereotactic radiosurgery and radiotherapy (SRS/SRT) programs. With this software, clinicians can rapidly create highly conformal treatment plans that use cones or multileaf collimators (MLCs) from a variety of manufacturers to help spare healthy tissue. A combination of clinical protocols, real-time dose updates, and optimization speed the planning process for both intracranial and extracranial SRS.

- Complete Stereotactic Solution
- High-Quality Radiosurgery Plans
- Fast Planning
- An Integrated Solution

Proton Treatment

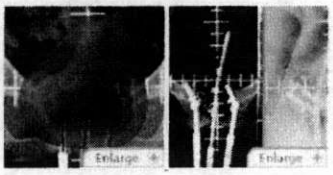


Eclipse Proton Planning combines the latest in fast, accurate proton calculation algorithms with the power of the Eclipse treatment planning system to create a single system for proton, photon, electron, and brachytherapy planning. Proton plans can be combined and compared with other modalities, or used as the base dose for a boost.

Eclipse supports a variety of proton beam lines as well as ocular planning. Clinicians use automated proton planning tools to easily take advantage of the unique physics of protons to reduce healthy tissue damage.

- State-of-the-Art Algorithms
- Powerful Contouring Tools
- Automatic Field Setup
- Multi-Modality Imaging
- Open System
- Integrated Environment
- Ocular Proton Planning

Brachytherapy



The combination of the BrachyVision™ 3D treatment planning system with the VariSource™ 200 HDR afterloader and GammaMedPlus™ HDR/PDR afterloader establishes Varian as a leader in brachytherapy.

BrachyVision works as a standalone brachytherapy planning system or as part of an Eclipse treatment planning system. For those customers who wish to perform film-based brachytherapy planning, Eclipse includes BrachyVision 2D. BrachyVision 3D is available as an option for Eclipse users who require image-guided 3D brachytherapy planning.

- Film-based Brachytherapy Planning
- Supported Sources and Techniques
- Flexible Plan Normalization
- Plan Analysis

Contact Varian Oncology

Varian Oncology Headquarters
Tel: 1.650.424.5700
Contact Us

Events

- **08/08/2010 - 08/13/2010**
  CAARI 21st International Conference on the Application of Accelerators in Research & Industry
- **08/08/2010 - 08/21/2010**
  15th Brazilian Congress of Medical Physics
- **08/11/2010 - 08/13/2010**
  MOGA - Medical Oncological Group of Australia
- **08/18/2010 - 08/21/2010**
  World Cancer Congress (International Union Against Cancer)

VARIAN medical systems | A partner for life

Oncology | Radiation Oncology
Eclipse | Planning Technology

- **ARIA**
  Information System
- **Eclipse**
  Treatment Planning
- Acuity
  Verification & Simulation
- **Trilogy**
  Treatment Delivery
- Clinac
  Treatment Delivery
- Upgrades

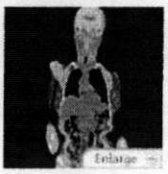
Eclipse | Planning Technology

Treatment planning with Eclipse™ is a fast process, easily accomplished throughout the department. With Eclipse, the radiation therapy team can efficiently plan and deliver the best course of treatment for each patient.

- Initial commissioning, a typically time-consuming task for the physics staff, is considerably faster and simpler with Eclipse.
- Virtual simulation, an integral part of Eclipse, eliminates data transfer steps and the management of separate sets of data.
- Emergency treatments can be planned immediately in the simulation area.
- Powerful tools simplify contouring tasks and field set up, and a fast calculation engine reduces treatment planning time.
- Using interactive optimization in Eclipse, a clinician can rapidly create high quality intensity-modulated radiation therapy (IMRT) plans.

Integrated plan verification and quality assurance tools speed ongoing treatment planning and beam data validation, saving time for physicists. For example, with the portal dosimetry option, the physics time required for IMRT pretreatment verification is reduced to less than 20 minutes per patient compared to hours with film. High-resolution IMRT treatments planned using Eclipse can typically be delivered in a normal 15-minute time slot on a Varian® Clinac® linear accelerator.

Virtual Simulation and Field Setup



Eclipse integrates virtual simulation and treatment planning in one comprehensive system. Virtual simulation and planning processes are accelerated through excellent segmentation software that includes:

- Comprehensive clinical protocol templates
- Automatic field positioning
- Hot-keys
- Automatic field aperture shaping with a flexible interactive user interface.

Eclipse can create 3D patient models from any DICOM 3.0-compliant image set, including CT, MR, and PET.

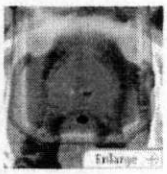
- Versatile Image Management
- Comprehensive Virtual Simulation
- Efficient Field Set up
- Unmatched Data and Process Integration

Contouring

For IMRT, 3D conformal, and proton therapy planning, segmenting organs at risk and accurately delineating target volumes are critical. The powerful contouring tools in Eclipse reduce structure segmentation time from hours to minutes. Clinicians can accurately define targets and organs at risk on fused multimodality images with advanced drawing and editing capabilities. Enhanced templates and powerful post-processing of structures accelerate the contouring process.

- Comprehensive Image Visualization
- Smart Segmentation™ – Truly Automatic Segmentation
- Varian IKOE knowledge - based Segmentation
- Efficient Segmentation
- Powerful Editing & Post-Processing

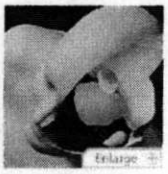
Dose Calculation



Advanced algorithms in Eclipse accurately and quickly calculate dose distributions for photons, electrons, protons, and brachytherapy. With the combination of modular algorithms and the flexible architecture of Eclipse, clinicians can select the optimum algorithm for each treatment modality. The clinician can rapidly customize IMRT plans using the interactive dose-volume optimization. A flexible calculation framework increases efficiency in the treatment planning process, particularly in a distributed planning environment. Specialized beam data analysis tools reduce the time for commissioning.

- Photons
- Electrons
- Protons
- Rapid Optimization

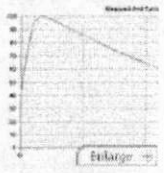
Plan Evaluation



An oncology team can decide on the most effective course of treatment for each patient by comparing different plan modalities. Using Eclipse, clinicians can combine, compare, and evaluate different candidate plans and even different treatment modalities on a single integrated system. When Eclipse workstations are distributed throughout a clinic, plans can be evaluated anywhere and anytime, greatly speeding the planning process.

- Multiple Dose Displays
- Powerful Plan Comparison
- Flexible Plan Documentation
- Electronic Plan Approval

Commissioning and QA



Designed to speed initial commissioning, Eclipse features treatment machine creation scripts, rapid beam data import, automated configuration, and integrated beam data analysis. System quality is easily maintained through data integrity checks and electronic beam data approval, and portal dosimetry revolutionizes pretreatment verification.

In addition, integration in the Varian environment helps simplify quality assurance for the planning process, and DICOM transfer to third-party systems.

- Rapid Beam Data Entry and Documentation
- Efficient Beam Model Validation
- Integrated Plan Verification and Quality Assurance

Contact Varian Oncology

Varian Oncology Headquarters
Tel: 1.650.424.5700
Contact Us

Events

- **08/08/2010 - 08/13/2010**
  CAARI 21st International Conference on the Application of Accelerators in Research & Industry
- **08/08/2010 - 08/21/2010**
  15th Brazilian Congress of Medical Physics
- **08/11/2010 - 08/13/2010**
  MOGA - Medical Oncological Group of Australia
- **08/18/2010 - 08/21/2010**
  World Cancer Congress (International Union Against Cancer)

VARIAN medical systems | A partner for life

Oncology | Radiation Oncology
Hardware Specifications

- **ARIA**
  Information System
- **Eclipse**
  Treatment Planning
- **Acuity**
  Verification & Simulation
- **Trilogy**
  Treatment Delivery
- **Clinac**
  Treatment Delivery
- Upgrades

Hardware Specifications

Varian requires that the Eclipse™ Treatment Planning System computer equipment and operating system software be purchased from Varian. When computer equipment and operating system software is purchased from Varian, it is warranted under the standard Varian maintenance contract.

See additional information for assistance in the following areas of department planning:

- Network Configuration Guidelines
- Departmental Needs and Requirements
- Eclipse Workstations
- Citrix Server
- High Availability and Rapid Recovery Protection (HARRP) Server
- Other System Requirements
- General Advice

Contact Varian Oncology

Varian Oncology Headquarters
Tel: 1.650.424.5700
Contact Us

Events

- **08/08/2010 - 08/13/2010**
  CAARI 21st International Conference on the Application of Accelerators in Research & Industry
- **08/08/2010 - 08/21/2010**
  15th Brazilian Congress of Medical Physics
- **08/11/2010 - 08/13/2010**
  MOGA - Medical Oncological Group of Australia
- **08/18/2010 - 08/21/2010**
  World Cancer Congress (International Union Against Cancer)

VARIAN medical systems | A partner for life

Oncology | Radiation Oncology
Eclipse | Benefits

- **ARIA**
  Information System
- **Eclipse**
  Treatment Planning
- Acuity
  Verification & Simulation
- **Trilogy**
  Treatment Delivery
- Clinac
  Treatment Delivery
- Upgrades

Eclipse | Benefits

Eclipse™ customers are talking about how Eclipse is so easy to learn and easy to use. They describe the flexibility of the latest techniques and ability to devise the best personalized plans for their patients in the least amount of time.

Customer quotes:

Randall W. Holt, Ph.D.
Director of Physics
North Valley Radiation Oncology

"One of the really nice things about Eclipse especially connected with Aria is that anything you do on Eclipse effectively becomes the plan and the treatment. There's virtually no reason for us to do any additional steps in order to connect the two together. So they're virtually one and the same."

Francine Sheridan
Chief Dosimetrist/Therapist
Dale and Frances Hughes Cancer Center

"The QA process within Eclipse is very easy and intuitive. It enables you to take a plan with one or two clicks, place it onto a phantom, do the calculation, and prepare it for the physicist."

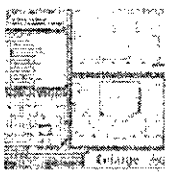
Gordon R. Ray, M.D.
Chairman, Department of Radiation Oncology
Palo Alto Medical Foundation

"The speed of the calculations in Eclipse IMRT is significantly improved. This greatly increased the throughput in our department, and that's translated into offering physicians a better choice of plans. And we think that's translated into better care."

Mickey Goldman, FACHE
Administrator
Goldman Consulting

"The efficiency of the QA process within Eclipse has been significant, particularly with IMRT plans. Where it used to take anywhere from an hour to an hour and a half, they are basically able to, from start to finish, complete that task in about thirty minutes."

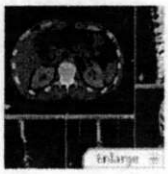
Physicist



Today's radiation therapy physicist encounters a multitude of technologies for managing the radiation dose delivered to patients. The physicist ensures the quality of the underlying beam models used for treatment planning, and verifies, approves, and controls the quality of the beam data. From initial commissioning to the routine quality assurance of treatment plans, the wide range of powerful tools and capabilities in Eclipse accelerates and simplifies these tasks for the physicist.

- Integrated and Efficient Process
- Advanced Treatment Planning and Delivery
- Rapid Commissioning
- Complete Quality Assurance

Dosimetrist

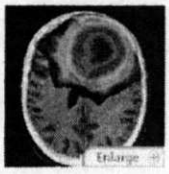


The adoption of 3D conformal and intensity-modulated radiation therapy (IMRT) for routine treatments has added to the heavy workloads of dosimetrists. Eclipse is a comprehensive treatment planning system that simplifies modern radiation therapy planning for all modalities, such as 3D conformal, IMRT, electron, proton, and brachytherapy.

Structure templates and automated structure segmentation in Eclipse reduce contouring time. Eclipse clinical protocol templates accelerate the planning process by automatically generating plan parameters from the physician's intent. Dosimetrists can easily create and compare variations of a plan to obtain the best treatment for the patient.

- Comprehensive Planning Solution
- Truly Automatic Contouring
- Fast Manual Contouring & Editing
- Simple Field Setup
- Insightful Plan Evaluation

Administrator



Managed healthcare costs continue to affect the delivery of radiation therapy even as more complex and effective treatment methods gain use in the radiation oncology clinic. Administrators face the challenge of improving the quality of care by implementing advanced techniques while maintaining cost effectiveness. With Eclipse, efficiency and cost effectiveness increase, providing true security of investment while improving quality of care.

- Efficiency by Design
- Comprehensive Capabilities
- Secure Investment
- Improved Quality of Care

Radiation Oncologist

Eclipse is a comprehensive treatment planning system that simplifies the complexity modern radiation therapy brings to radiation oncologists. Extensive templates facilitate standards of care and protocols. Yet the flexibility of Eclipse allows the clinician to rapidly tailor plans for each patient.

Advanced contouring and plan evaluation tools, combined with templates and interactive optimization, speed the planning process for complex 3D, IMRT and image-guided radiation therapy (IGRT) treatments. With this level of integration, the physician can plan treatments from anywhere in the department.

- Clinical Flexibility
- Complete IGRT Support
- Extensive Clinical Protocol Support
- Rapid Contouring
- Comprehensive Plan Evaluation
- Convenient Access

Contact Varian Oncology

Varian Oncology Headquarters
Tel: 1.650.424.5700
Contact Us

Events

- **08/08/2010 - 08/13/2010**
  CAARI 21st International Conference on the Application of Accelerators in Research & Industry
- **08/08/2010 - 08/21/2010**
  15th Brazilian Congress of Medical Physics
- **08/11/2010 - 08/13/2010**
  MOGA - Medical Oncological Group of Australia
- **08/18/2010 - 08/21/2010**
  World Cancer Congress (International Union Against Cancer)

- Annual Reports
- Stock Information
- Financials
- Events and Webcasts
- SEC Filings
- XBRL Filings
- Corporate Governance
- Shareholder Services
- Investor Contacts

VARIAN medical systems | A partner for life

Oncology | Radiation Oncology
Eclipse | Product Resources

- **ARIA**
  Information System
- **Eclipse**
  Treatment Planning
- Acuity
  Verification & Simulation
- **Trilogy**
  Treatment Delivery
- Clinac
  Treatment Delivery
- Upgrades

Eclipse | Product Resources

In support of Varian's Eclipse™ Community, this section provides links to Eclipse user groups, our product support helpdesk, product services, hardware specifications, and a variety of policy and compliance statements.

Please contact your local sales representative to learn more about Eclipse.

- MyVarian
- Eclipse Support
- Brochures
- Download
- Product Services
- Policies and Statements
- Hardware Specifications

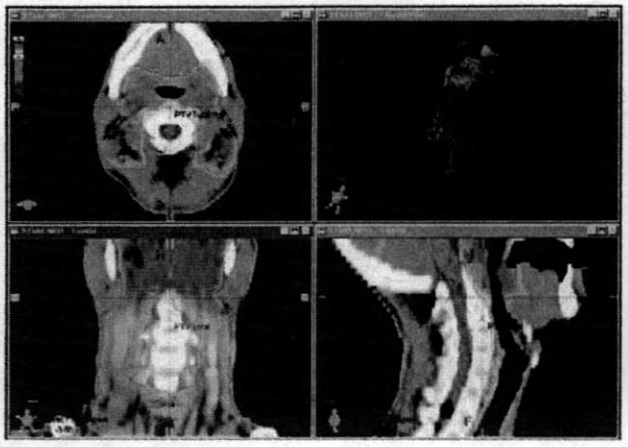
Contact Varian Oncology

Varian Oncology Headquarters
Tel: 1.650.424.5700
Contact Us

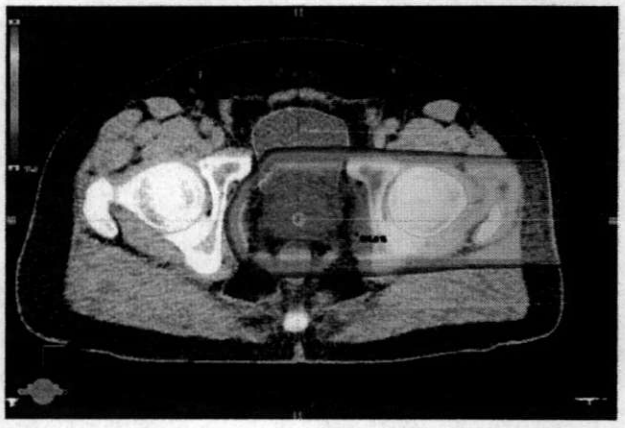
Events

- **08/08/2010 - 08/13/2010**
  CAARI 21st International Conference on the Application of Accelerators in Research & Industry
- **08/08/2010 - 08/21/2010**
  15th Brazilian Congress of Medical Physics
- **08/11/2010 - 08/13/2010**
  MOGA - Medical Oncological Group of Australia
- **08/18/2010 - 08/21/2010**
  World Cancer Congress (International Union Against Cancer)

Eclipse™ Treatment Planning System



A dose color wash for a head and neck IMRT plan shows excellent sparing of the spinal cord and the parotid glands.



A single proton beam treats the prostate with a highly conformal dose, with avoidance of critical structures and minimal integral dose to the body.

Flexible Plan Evaluation

An oncology team can decide on the most effective course of treatment for each patient by comparing different plan modalities. Clinicians can combine, compare, and evaluate different treatment modalities on a single integrated Eclipse planning system. They can also conveniently compare candidate plans. Eclipse workstations can be distributed throughout a clinic so plans can be evaluated anywhere and at anytime, which speeds the planning process.

Easy Commissioning and QA

Eclipse simplifies and accelerates the entire treatment planning process. Treatment machine creation scripts, rapid beam data import, automated configuration, and integrated beam data analysis speed initial commissioning.

System quality is easily maintained through data integrity checks and electronic beam data approval. Pre-treatment verification has been revolutionized with portal dosimetry. QA of the planning process is simplified through integration in the Varian Inspiration environment and DICOM transfer to third-party systems.

Effective Teamwork

Treatment planning with Eclipse is a process which is distributed efficiently throughout the department. Physicians can delineate targets quickly on multi-modality images at the simulator or in their offices. Dosimetrists can rapidly generate high-quality plans that are easily verified by physicians. Therapists can deliver complex IMRT treatments on a Clinac® linear accelerator in a typical 15-minute time slot. Data transfer is eliminated from the treatment planning process because patient data is centralized and available anywhere in the department.

Improved Quality of Care

With easy commissioning, protocol support, fast planning, and a consistent user interface, a clinic can readily implement IMRT, IGRT, and DART to deliver the most effective treatments with the fewest complications. Because Eclipse is comprehensive, the radiation therapy team can efficiently plan and deliver the best course of treatment for each patient.

VARIAN medical systems | ONCOLOGY SYSTEMS

USA Headquarters, California
Varian Medical Systems
Palo Alto, CA
Tel: 650.424.5700
800.544.4636
Fax: 650.424.5358
www.varian.com

Headquarters Europe, Eastern Europe, Africa, Middle & Near East
Varian Medical Systems International AG
Zug, Switzerland
Tel: 41.41.749.8844
Fax: 41.41.749.8844
info.europe@varian.com

RAD7158B

7/06 (5K)

Eclipse™ Treatment Planning System

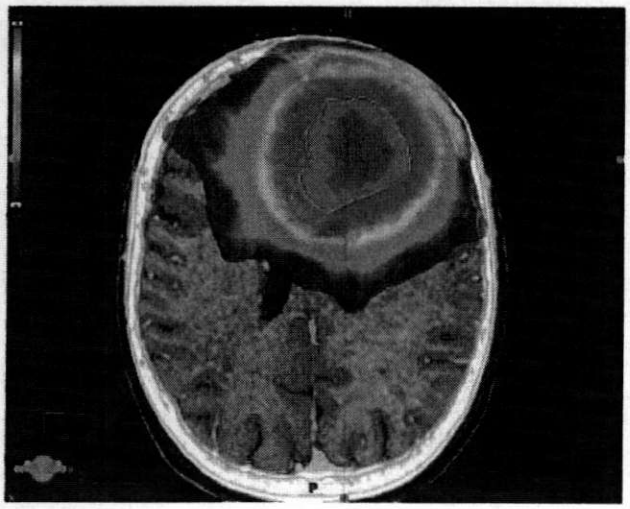
Administrator

Managed healthcare costs continue to affect the delivery of radiation therapy even as more complex and effective treatment methods requiring advanced techniques gain use in the radiation oncology clinic. Administrators face the challenge of improving the quality of care by implementing advanced techniques while maintaining cost effectiveness. With Eclipse, efficiency and cost effectiveness for advanced techniques increase, providing true security of investment while improving quality of care.

Efficiency by Design

By introducing Eclipse into the clinic, administrators can increase efficiency throughout the therapy process. Initial commissioning, a typically time-consuming task for the physics staff, is considerably faster and simpler with Eclipse. Virtual simulation, an integral part of Eclipse, eliminates data transfer steps and the management of separate sets of data. Emergency treatments can be planned immediately in the simulation area. Powerful tools simplify contouring tasks and field setup, and a fast calculation engine reduces treatment planning time. Using interactive optimization in Eclipse, a clinician can rapidly create high quality intensity-modulated radiation therapy (IMRT) plans.

Integrated plan verification and quality assurance tools speed ongoing treatment planning and beam data validation, saving time for physicists. For example, with the portal dosimetry option, the physics time required for IMRT pre-treatment verification is reduced to less than 20 minutes per patient compared to hours with film. High-resolution IMRT treatments planned using Eclipse can typically be delivered in a normal 15-minute time slot on a Varian Clinac® linear accelerator.



A dose distribution is displayed on fused CT and MR images.

Eclipse has comprehensive clinical protocol templates that speed the planning process. These protocols pre-populate all of the planning parameters based upon treatment technique, patient anatomy, and physician preferences. With clinical protocol templates, clinicians can create class solutions that can be easily individualized for the needs of each patient through interactive optimization. Easy transfer of these protocols between institutions simplifies the sharing of clinical experience and standardization of care.

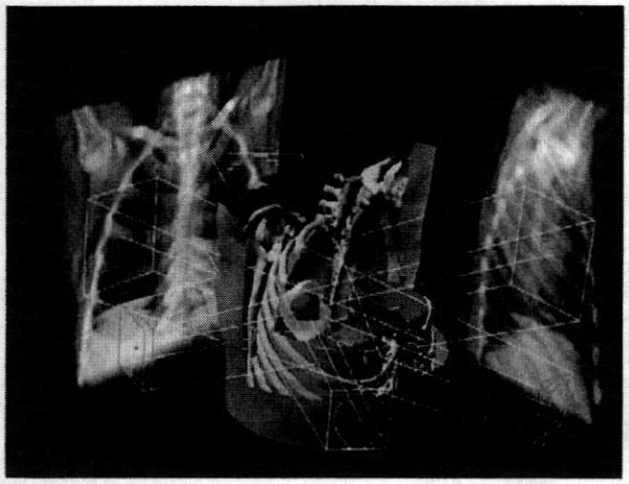
The clinical staff can access treatment planning data directly from anywhere in the clinic because patient data is centralized. When administrators introduce Varian's Inspiration™ integrated oncology environment into the clinic, all planning, verification, and treatment tasks are integrated, further improving efficiency in the radiation therapy process. For example, plans that are approved in Eclipse are immediately available for treatment at the accelerator and are an integral part of the patient record in the ARIA™ oncology information system.

Comprehensive Capabilities

All conventional techniques including 2D, 3D conformal, electron, and brachytherapy can be planned on Eclipse. The administrator can confidently implement the most advanced treatment techniques including IMRT, stereotactic radiosurgery, respiratory gating, and proton therapy with treatment equipment from all major vendors in radiation therapy. Because Eclipse is comprehensive, oncologists can combine and compare plans for different treatment modalities to determine the best course for a patient.

Eclipse supports image-guided radiation therapy (IGRT) and DART™ dynamic adaptive radiation therapy to improve patient treatment. Image fusion of multi-modality images helps the clinician to clearly see targets and critical structures. Eclipse can reconstruct dose distributions based upon cone-beam CT images and actual treatment parameters for a true 3D treatment record. This information helps the clinician to adapt the treatment to changing conditions of the patient. To account for organ motion during respiration, Eclipse supports the import of 4D images.

Eclipse™ Treatment Planning System



Patient anatomy, digitally reconstructed radiographs (DRRs), and field arrangements can be rotated in 3D.

Secure Investment

Eclipse excels as a secure and cost-effective investment in a number of key strategic areas. With Varian's commitment and ongoing investment in treatment planning, an administrator can feel confident in purchasing not only the best planning system for today, but also a solid platform for the future of radiation oncology.

Eclipse is a well-supported, expandable system with a low cost of ownership. PC technology helps reduce not only hardware costs for both initial and additional Eclipse systems but also training costs. The centralized patient database and distributed architecture reduce the costs of maintenance and backups, and simplify adding more treatment planning, virtual simulation, and physician review workstations. Sharing patient information and departmental processes increases staff flexibility throughout the clinic and across multiple sites. For clinicians needing to work away from the clinic, complete Eclipse capability is available via high-speed internet connection using secure Eclipse RemoteConnect.

Eclipse is an open system that conforms to DICOM RT for communication with diagnostic imaging devices, CT simulators, information systems, treatment planning systems, and other third-party devices. As a clinic incorporates new devices, they can readily communicate with Eclipse.

Varian offers extensive support through its Clinical Support Services. Application support for Eclipse through Varian's extensive Help Desk infrastructure gets excellent reviews on customer satisfaction surveys. Remote installations, diagnostics, and service via SmartConnect℠ technology ensure rapid response, minimizing downtime and maximizing productivity. Varian also provides continuing education with CE credits to help keep the clinical staff knowledgeable and up to date through application training and other courses, as well as extensive electronic and on-line documentation.

Improved Quality of Care

Implementing Eclipse for treatment planning helps administrators achieve the goal of high-quality patient care. Class solutions and clinical protocols, which have been developed by oncologists, can be repetitively applied and easily shared between facilities using Eclipse clinical protocol templates. For clinics participating in clinical trials, Eclipse complies with RTOG electronic submission requirements for conventional, IMRT, and brachytherapy trials.

To ensure the integrity of patient data, a system of user rights restricts levels of access and approval for users. Eclipse has plan approval with electronic signatures and supports clinic requirements for HIPAA compliance.

For the best course of treatment for a patient, clinicians can create and compare plans using different treatment modalities. With easy commissioning, protocol support, fast planning, and a consistent user interface, clinics can easily implement advanced capabilities, such as IMRT, IGRT, and DART to deliver the most radiation to the tumor with the fewest complications.

VARIAN medical systems | ONCOLOGY SYSTEMS

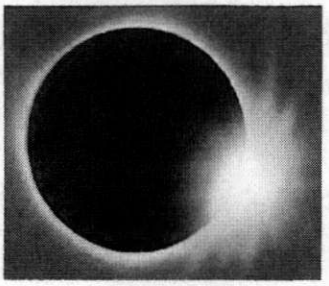
USA Headquarters, California
Varian Medical Systems
Palo Alto, CA
Tel: 650.424.5700
800.544.4636
Fax: 650.424.5520
www.varian.com

Headquarters Europe, Eastern Europe, Africa, Middle & Near East
Varian Medical Systems International AG
Zug, Switzerland
Tel: 41.41.749.8844
Fax: 41.41.749.8845
info.europe@varian.com

RAD7157B

7/06 (5K)

 2005



Eclipse™ Treatment Planning System

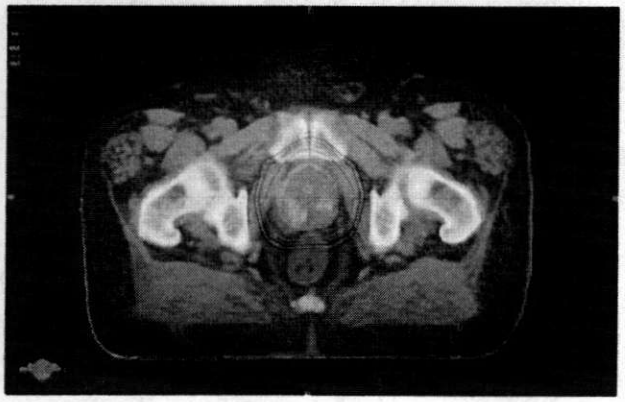
Overview

Eclipse is a comprehensive system that simplifies the complexity of modern radiation therapy planning for all modalities such as 3D conformal, intensity-modulated radiation therapy (IMRT), electron, proton, and brachytherapy. Advanced processes such as image-guided radiation therapy (IGRT) and DART™ dynamic adaptive radiation therapy are supported. Eclipse helps dosimetrists, physicists, and physicians efficiently create, select, and verify the best treatment plans for their patients. Automation facilitates standards of care and protocols; yet, with the flexibility of Eclipse, the clinician can rapidly tailor plans for each patient. The efficiency and cost effectiveness of Eclipse enables the adoption of advanced techniques, providing true security of investment while improving the quality of care.

Comprehensive Planning

Eclipse is a comprehensive system for planning all photon, electron, proton, and brachytherapy treatment modalities. This reduces staff training costs, enables multi-modality treatments, and facilitates the process for all types of planning, from simple IRREG plans calculated at the CT simulator to complex IMRT plans that require a team approach. Eclipse supports IMRT delivery modes of linear accelerators from all major vendors, including sliding window and step-and-shoot techniques. Varian offers a complete and comprehensive stereotactic planning solution for all accelerator-based stereotactic programs. Clinicians can plan with 4D image sets and fused functional and anatomical images for efficient IGRT.

In Varian's Inspiration™ integrated oncology environment, all parameters required for patient treatment are immediately available to the electronic medical record and treatment delivery system. All patient information is kept in one database, eliminating the need to export and import data from one application to another. Eclipse can reconstruct dose distributions based upon cone-beam CT (CBCT) images and actual treatment parameters for DART.



Cone-beam CT images can be fused and blended with the original CT images and used for DART.

Integrated Virtual Simulation

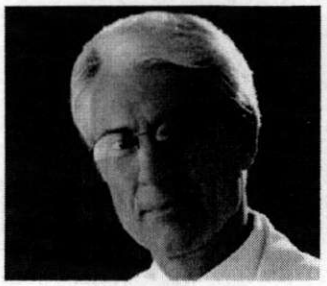
With Eclipse, virtual simulation and treatment planning are integrated in one comprehensive system. Virtual simulation and planning processes are accelerated through excellent segmentation software with comprehensive clinical protocol templates, automatic field positioning, hot-keys, and automatic field aperture shaping with a flexible interactive user interface. Eclipse can create 3D patient models from any DICOM 3.0-compliant image set, including CT, MR, and PET.

Rapid Contouring

For IMRT, 3D conformal, and proton therapy planning, segmenting organs at risk and accurately delineating target volumes are critical. Structure segmentation time is reduced from hours to minutes with the powerful contouring tools in Eclipse. Clinicians can accurately define targets and organs at risk on fused multi-modality images with advanced drawing and editing capabilities. Enhanced templates and powerful post-processing of structures accelerate the contouring process.

Accurate Dose Calculation and Optimization

Advanced algorithms in Eclipse accurately and quickly calculate dose distributions for photons, electrons, protons, and brachytherapy. With the combination of modular algorithms and the flexible architecture of Eclipse, clinicians can select the optimum algorithm for each treatment modality. The clinician can rapidly customize IMRT plans using the interactive dose-volume optimization. A flexible calculation framework increases efficiency in the treatment planning process, particularly in a distributed planning environment. Specialized beam data analysis tools reduce the time for commissioning.



Eclipse™ Treatment Planning System

Physicist

Today's radiation therapy physicist encounters a multitude of technologies for managing the radiation dose delivered to patients. The physicist ensures the quality of the underlying beam models used for treatment planning, and verifies, approves, and controls the quality of the beam data. From initial commissioning to the routine quality assurance of treatment plans, the wide range of powerful tools and capabilities in Eclipse speeds these tasks for the physicist.

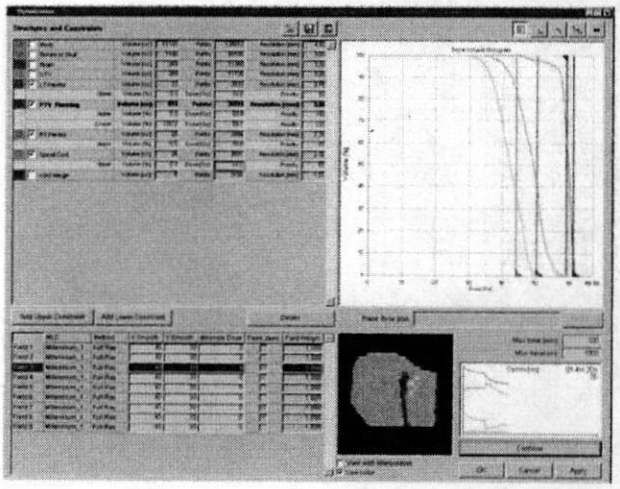
Integrated and Efficient Process

Eclipse is a comprehensive system for planning all photon, electron, brachytherapy, and proton treatment modalities. In Varian's Inspiration™ integrated oncology environment, all parameters required for patient treatment are immediately available to the electronic medical record and treatment delivery systems without importing or exporting. With extensive support of the DICOM standard, treatment plan information is available to multiple vendor devices.

Advanced Treatment Planning and Delivery

With Eclipse, the physicist can confidently implement the most advanced treatment techniques and processes in radiation therapy, including intensity modulated radiation therapy (IMRT), image-guided radiation therapy (IGRT), DART™ dynamic adaptive radiation therapy, and respiratory gating. IMRT plans are optimized for both dose-volume objectives and delivery efficiency. Dose-volume optimization in Eclipse is interactive, so the physicist can quickly achieve clinical objectives with excellent agreement between optimal and delivered dose distributions. Eclipse clinical protocol templates increase IMRT planning efficiency and clinic-wide standardization of treatment planning techniques by tying together prescription information, plan templates, structure lists, optimization criteria, and plan review preferences in a single location.

The seventh-generation Leaf Motion Calculator (LMC) takes full advantage of the multileaf collimators from several accelerator vendors. The LMC minimizes wear and tear on the multileaf collimator and maximizes the efficiency of IMRT treatment delivery whether employing sliding window or segmented delivery techniques.



Eclipse updates optimization information with each iteration so the clinician can interactively modify parameters based on real-time feedback.

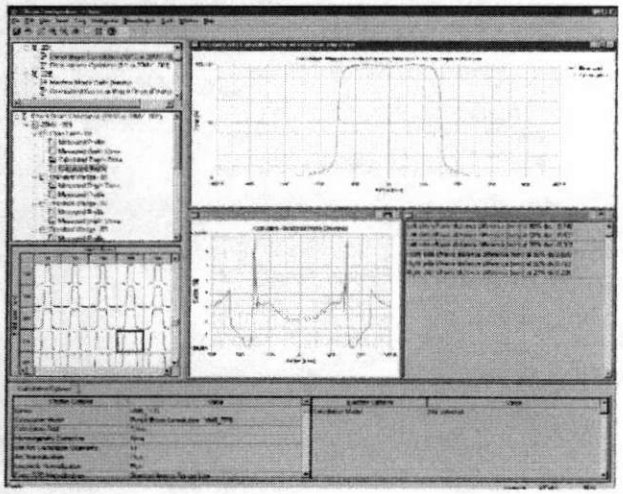
In support of IGRT, multiple image modalities such as CT, PET, and MR can be fused automatically or manually. Co-registration of 4D CT and CT/PET image sets in Eclipse supports motion management and respiratory gating in treatment planning. Physicists can also use cone-beam CT (for example, from the Acuity™ planning, simulation and verification system, and Varian's On-Board Imager® kV imaging system) and other post-planning image sets for adaptive planning during the course of treatment.

Eclipse™ Treatment Planning System

Rapid Commissioning

Eclipse simplifies beam configuration and commissioning for the physicist. Configuration of the calculation model is automated so that the entire process takes as little as 30 minutes once the data set is read into the system. As recommended by the AAPM Task Group 53 on Quality Assurance for Clinical Radiotherapy Treatment Planning, Eclipse calculates the dose distribution to a water phantom using the configured beam models. The calculated data are then automatically compared with the measured data.

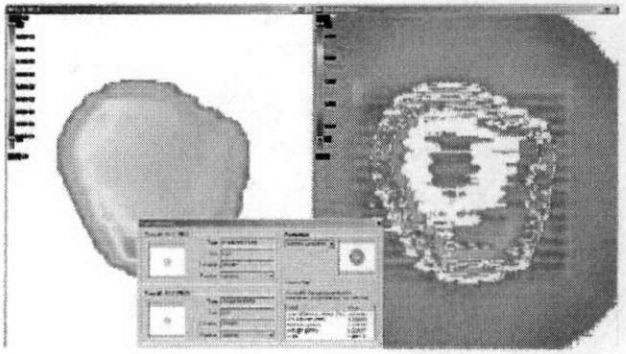
For rapid and simplified commissioning, sets of pre-measured beam data are available for the Trilogy™ system, Clinac® iX, EX, and most C/CD high-energy linear accelerators. Beam data from a configured machine can be printed as well as exported in tabular format to most popular spreadsheet applications for off-line comparisons.



Measured beam data is compared to calculated beam data.

Complete Quality Assurance

The integrity of commissioned beam data and configured beam models is protected against unauthorized modification by requiring password approval of all beam data before it can be used clinically. The use of unapproved data in dose calculations is thereby prevented. Warnings are issued to the user when attempting to recalculate plans using modified data.



Calculated and measured portal dose distributions are compared for pre-treatment verification using gamma analysis.

Quality assurance of dose distributions is simplified by verification plans that can be automatically created in Eclipse. The physicist can calculate a patient's plan in a scanned or geometric phantom for comparison measurements. Alternatively, the Portal Dose Calculation algorithm calculates the dose distribution measured by the PortalVision™ MV imaging system at the imaging plane for pre-treatment QA, eliminating the need for film.

Verification plans can be transferred to the linear accelerator for easy quality assurance. The verification plan dose distribution and all plan parameters, including IMRT fluence, can be exported to DICOM RT-compatible quality assurance systems for further analysis. For physicists needing to work away from the clinic, complete Eclipse capability is available via high-speed internet connection using secure Eclipse RemoteConnect solution.

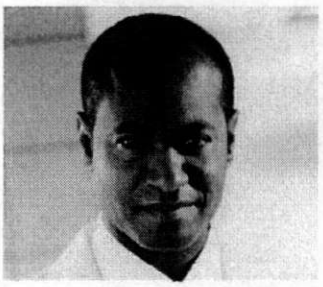
VARIAN medical systems | ONCOLOGY SYSTEMS

USA Headquarters, California
Varian Medical Systems
Palo Alto, CA
Tel: 650.424.5700
800.544.4636
Fax: 650.424.5973
www.varian.com

Headquarters Europe, Eastern Europe, Africa, Middle & Near East
Varian Medical Systems International AG
Zug, Switzerland
Tel: 41.41.749.8844
Fax: 41.41.749.8840
info.europe@varian.com

RAD7142B

7/06 (5K)



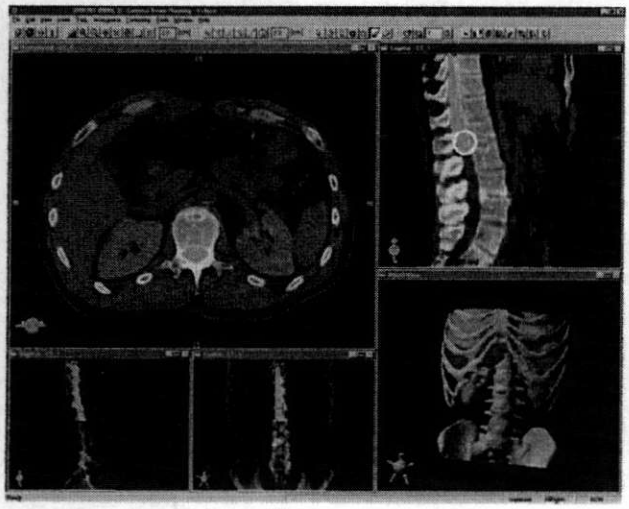
Eclipse™ Treatment Planning System

Dosimetrist

The adoption of 3D conformal and intensity-modulated radiation therapy (IMRT) for routine treatments has added to the heavy workloads of dosimetrists. Eclipse is a comprehensive treatment planning system that simplifies the complexity of modern radiation therapy planning for all modalities, such as 3D conformal, IMRT, electron, proton, and brachytherapy. Structure templates and automated structure segmentation in Eclipse reduce contouring time. Eclipse clinical protocol templates accelerate the planning process by automatically generating plan parameters from the physician's intent. Dosimetrists can easily create and compare variations of a plan to obtain the best treatment for the patient.

Comprehensive Planning Solution

To give the best possible treatment, dosimetrists can choose and combine all types of radiation therapy, including photon (conventional, arc, and IMRT), electron, proton, and brachytherapy. Eclipse simplifies the use of advanced planning techniques such as field-in-field, electronic surface compensation, and IMRT. The clinician only needs to learn one system for all treatment planning needs.



An intelligent 3D paintball tool quickly and accurately contours the spinal cord on a sagittal plane.

User-defined clinical protocol templates automate every step of the planning process based upon disease stage, clinician's treatment preference, or RTOG study. Clinical protocol templates contain structure sets, field geometry, accessories, dose prescription, treatment plan objectives, IMRT objectives, and dose-volume histogram (DVH) calculation settings. Any plan may be saved as a clinical protocol, organized into user-defined folders, and applied consistently to subsequent patients. Eclipse sets up the planning parameters so the dosimetrist can focus on getting the best plans.

Eclipse stores patient information in a central database and makes it available at any workstation. Physicians can define targets and review plans from their offices instead of tying up systems in dosimetry. Virtual simulation is an integral part of the distributed planning environment. A comprehensive system of user rights includes electronic approval that locks plans from unintended modification. In Varian's Inspiration™ integrated oncology environment, all parameters required for patient treatment are immediately available to the electronic medical record and treatment delivery system.

Fast Contouring

Dosimetrists spend much of their time delineating target volumes and organs at risk. With Eclipse, contouring time is measured in minutes, not hours.

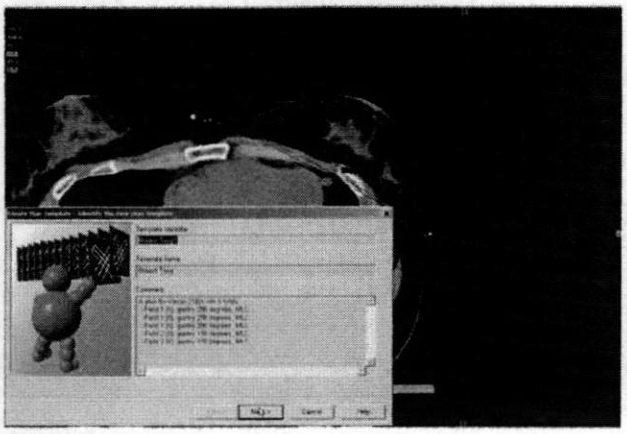
To speed the contouring process, Eclipse automatically creates structure sets from an anatomical database that includes structure names, display properties, ICD codes, volume types, and automated segmentation parameters.

A dosimetrist can quickly and easily register DICOM-compliant images such as CT, MR, and PET, and contour on blended image sets. Multiple contrasting color schemes highlight the tumor and critical structures in blended images. A slider bar makes it easy to adjust the relative weight of each image.

Multiple interactive layouts in Eclipse increase contouring efficiency and accuracy. The dosimetrist can easily zoom, pan, and scroll through image data using the mouse wheel. Drawing on sagittal and coronal views reveals structure details not seen on transverse views. All structures and organs are rendered in 3D to visually verify the segmentation.

2057

Eclipse™ Treatment Planning System



Powerful plan templates speed field setup by eliminating the need to manually enter plan parameters.

Automatic contouring tools accelerate the segmentation process. Many common structures can be contoured with a single mouse click. An adaptive paintbrush uses edge-detection algorithms to morph the shape being drawn for rapid and accurate segmentation. The dosimetrist can draw, edit, and stretch structures all with a single freehand tool. Creation of new structures from existing structures takes seconds using Boolean logic expressions in Eclipse. For instance, a wall of a given thickness can be created from an existing structure. Advanced editing and post-processing functions streamline structure modification and clean-up.

Simple Field Setup

Clinical protocol templates, graphical field design, hot keys, and direct parameter entry simplify field setup in Eclipse. Fields can be defined graphically with 2D planar, 3D, or beam's eye views (BEVs). Hot-keys reduce multiple-step field-positioning tasks to a single button click. For example, pushing a single button reshapes all field apertures to fit the target volume. Field alignment rules simplify the creation of complex field arrangements. With the rules activated, the user can simply move one field, and Eclipse will automatically reorient the other fields. Background calculation lets the dosimetrist work on other plans during dose calculation.

The planner can adjust field parameters while viewing dose distribution to quickly improve plans. Eclipse automatically converts an isodose surface into a structure. The simple stroke of a brush creates skin flash or edits fluence to remove hot or cold spots. In field-in-field planning, the dosimetrist can use isodose levels to rapidly create subfields that Eclipse can combine into a single field for treatment.

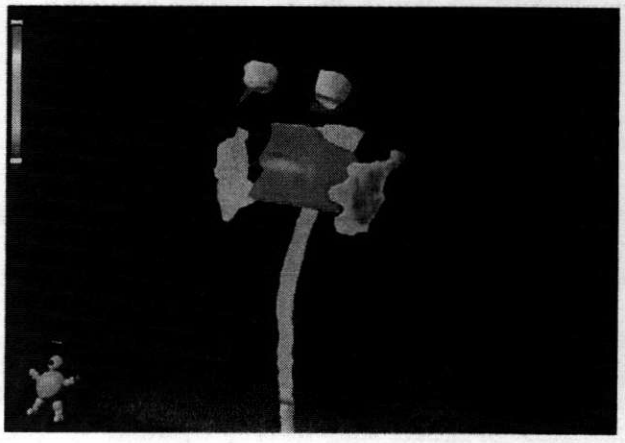
For IMRT, the planner can effectively apply clinical judgment in real-time by adjusting the DVH criteria while the optimization engine is running. Eclipse quickly shows the trade off between sparing and target coverage. A normal tissue constraint gives highly conformal plans without the creation of pseudo-structures for optimization.

Insightful Plan Evaluation

In Eclipse, dosimetrists can combine, compare, and evaluate all treatment planning modalities in a single integrated planning system. Determining the optimal treatment plan is easy through flexible screen layouts and plan comparison tools. Locked viewing of side-by-side plans helps the planner to compare alternative dose distributions. Eclipse displays axial views of up to six plans at one time. Plan summation and subtraction highlight the strengths and weaknesses of each plan.

Flexible DVH viewing reveals the differences between plans. The user can select any point along a DVH curve and the appropriate isodose level will display on the orthogonal plan views. Boolean logic operators show the DVH of combined structures without having to create the structures. A clinician can compare multiple plans on the same DVH and look at DVH difference plots.

The dosimetrist can generate documentation that meets the clinic's unique requirements. Hard-copy reports include BEV plots, isodose plots, DVHs, and screen printouts. Dosimetrists can create customized plan reports using a simple text editor. Screen captures can be incorporated into presentations and technical papers.



Eclipse maps the 3D dose to the surfaces of the target and critical structures.

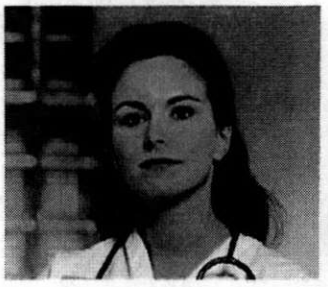
VARIAN medical systems | ONCOLOGY SYSTEMS

USA Headquarters, California
Varian Medical Systems
Palo Alto, CA
Tel: 650.424.5700
800.544.4636
Fax: 650.424.5358
www.varian.com

Headquarters Europe, Eastern Europe, Africa, Middle & Near East
Varian Medical Systems International AG
Zug, Switzerland
Tel: 41.41.749.8844
Fax: 41.41.749.8844
info.europe@varian.com

RAD7141B

7/06 (5K)



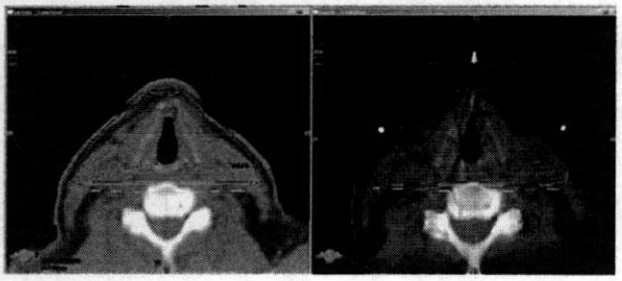
Eclipse™ Treatment Planning System

Radiation Oncologist

Eclipse is a comprehensive treatment planning system that simplifies the complexity modern radiation therapy brings to radiation oncologists. Extensive templates facilitate standards of care and protocols; yet, with the flexibility of Eclipse, the clinician can rapidly tailor plans for each patient. Advanced contouring and plan evaluation tools, combined with templates and interactive optimization, speed the planning process for complex 3D, intensity-modulated radiation therapy (IMRT) and image-guided radiation therapy (IGRT) treatments. With integration, the physician can plan treatments from anywhere in the department.

Clinical Flexibility

With Eclipse, oncologists can rapidly customize treatment plans for the needs of each patient. The oncologist chooses from a broad range of treatment options. All types of radiation therapy delivery are supported including photon (conventional and IMRT), electron, proton, and brachytherapy.



Larynx plans calculated on cone-beam CT images can be compared to the initial plan and used later for DART.

The IMRT delivery modes of accelerators from all major vendors are supported in Eclipse, including sliding window and step-and-shoot techniques. IMRT optimization is interactive, bringing clinical judgment and speed to the optimization process. The oncologist adjusts treatment objectives during optimization and gets immediate feedback on the dose to the target and other structures. With Eclipse, the oncologist can plan IMRT treatments using multiple targets with a single isocenter, non-coplanar beams, and fixed jaws for highly conformal dose distributions with efficient delivery. Base dose planning takes previously delivered dose distributions into account during optimization.

Planning for even the most complex cases is simplified in Eclipse. Several potential plans, even employing different modes of delivery, can be easily compared to choose the most suitable treatment plan for the patient.

Complete IGRT Support

Recent advances in radiation therapy are incorporated into Eclipse to support IGRT. Fusion of CT, MR, and PET images enhances functional and structural information while contouring or reviewing plans. The oncologist can use 4D image sets to account for organ motion in treatment plans.

Cone-beam CT (CBCT) images taken on the day of treatment can be used to modify the treatment plan based on changes in patient anatomy and physiology. Dosimetrists can recalculate dose on CBCT images registered to the planning images so that the actual dose delivered to the patient can be evaluated. In an example of the process known as DART™ dynamic adaptive radiation therapy, the plan can be re-optimized for future fractions to account for changes.

Extensive Clinical Protocol Support

Oncologists develop clinical protocols and class solutions in Eclipse using extensive customizable templates for plans, structures, and clinical objectives. Eclipse clinical protocol templates tie together prescription information, plan setups, structure lists, and plan review preferences in a single location automating much of the treatment planning process. These templates are specific to treatment technique, patient anatomy, and physician preference and promote standardized quality of care. Using templates as a starting point, the clinician can tailor the best possible treatment plan for the patient.

Clinical trials are important in furthering the effectiveness of radiation therapy. The clinician can build templates in Eclipse conforming to the treatment techniques and dose objectives specified by clinical trials. Eclipse complies with RTOG electronic submission requirements so the clinician can easily participate in conventional, IMRT, and brachytherapy trials.

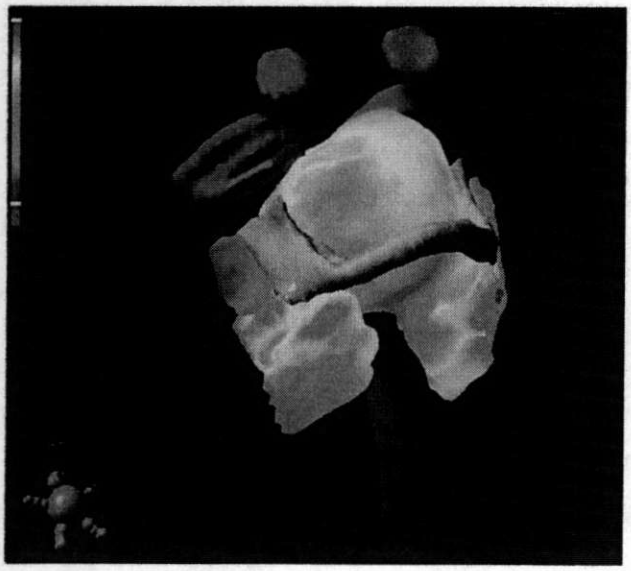
Eclipse™ Treatment Planning System

Rapid Contouring

Easy-to-learn contouring tools built into Eclipse simplify the complicated segmentation problems posed by current conventional and IMRT plans. A wide range of automatic, semi-automatic, and manual tools match any drawing preference. Intelligence is built into even the simplest tools so the oncologist spends less time meticulously outlining each organ of interest. Eclipse improves efficiency by combining drawing, editing, optimizing, panning, zooming, and scrolling in a unified mouse operating mode. For different perspectives, the oncologist can contour in coronal and sagittal as well as axial planes.

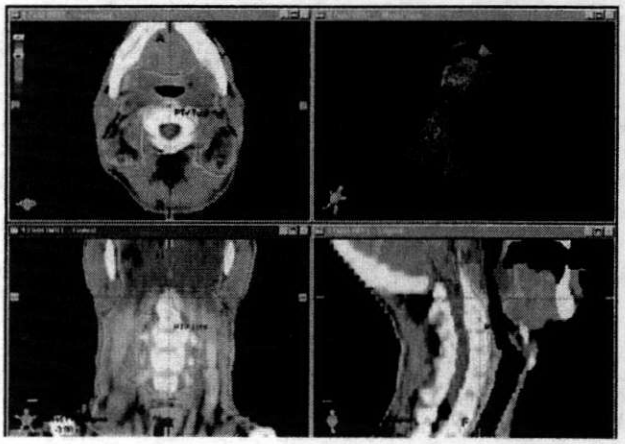
Comprehensive Plan Evaluation

With flexible dose displays and plan comparisons in Eclipse, the clinician can quickly and easily evaluate a large number of complex plans. User-defined isodose templates enhance isodose displays. Eclipse calculates dose-volume histograms (DVHs) with customizable line style and color, and displays hot spots on each slice and in the entire volume. For clinicians to easily visualize the spatial distribution of a plan, a 3D view shows isodose surfaces, the dose on the surface of each structure, and the volumetric hot spot. DVH curves are referenced to the isodose displays by drawing isodose curves corresponding to a selected point on the curve, thereby visually relating volumetric and spatial dose information.



With surface rendering, oncologists can easily evaluate the dose distribution of a plan.

Candidate plans can be displayed side-by-side and synchronized so that the same view from each plan is used in the evaluation. Eclipse can display DVH calculations from several plans on a single graph. Plans can be added or subtracted for rapid comparison of volumetric dose distributions.



Dose color wash for a head and neck IMRT plan showing excellent sparing of the spinal cord and the parotid glands.

Once approved, a plan cannot be modified, ensuring the integrity of the plan. Electronic plan treatment approval is password-protected. When Eclipse is part of the Varian Inspiration™ integrated oncology environment, the approved plan becomes immediately available for treatment without further data transfer. Therefore, the oncologist is assured that the patient will be treated as intended.

Convenient Access

Eclipse has a central database for patient information. All of the Eclipse features are accessible from any Eclipse computer in the department, including the oncologist's desktop. For example, prescription information and contours outlined by the oncologist are immediately available to the dosimetry staff without transfer. In the Inspiration environment, the oncologist can also review portal images, images from the On-Board Imager® kV imaging system, and patient setup statistics on the same system. For physicians needing to work away from the clinic, complete Eclipse capability is available via high-speed internet connection using secure Eclipse RemoteConnect.

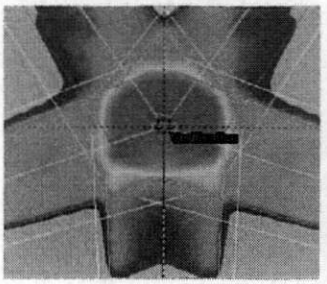
VARIAN medical systems | ONCOLOGY SYSTEMS

USA Headquarters, California
Varian Medical Systems
Palo Alto, CA
Tel: 650.424.5700
800.544.4636
Fax: 650.493.5634
www.varian.com

Headquarters Europe, Eastern Europe, Africa, Middle & Near East
Varian Medical Systems International AG
Zug, Switzerland
Tel: 41.41.749.8844
Fax: 41.41.740.5340
info.europe@varian.com

RAD7143B

7/06 (5K)



Eclipse™ Treatment Planning System

Commissioning and QA

Eclipse simplifies and accelerates the entire treatment planning process. Treatment machine creation scripts, rapid beam data import, automated configuration, and integrated beam data analysis speed initial commissioning. System quality is easily maintained through data integrity checks and electronic beam data approval. Pre-treatment verification has been revolutionized with Eclipse Portal Dose Calculation. Planning process QA is simplified through integration in the Varian Inspiration™ integrated oncology environment and DICOM transfer to third-party systems.

Rapid Beam Data Entry and Documentation

The initial configuration of any treatment planning system involves creation of treatment machines. Machine creation scripts, available in Eclipse for multiple vendors, simplify this process by eliminating manual entry of parameters and limits.

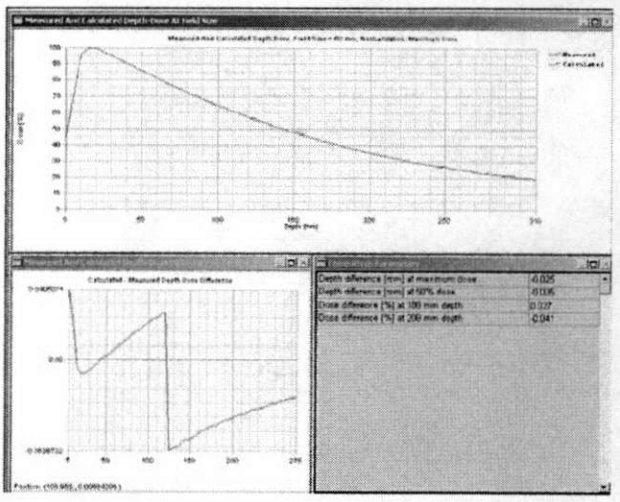
After beam data is imported into Eclipse and checked for consistency, configuration takes only a few minutes so the clinician can quickly begin validating the beam models. This may be simplified further by using Eclipse Beam Data for Clinac® linear accelerators.

The clinician can share or copy beam models between matched treatment units, making beam data management easier. With centralized beam data, the physicist only needs to maintain one set. Beam data can be printed in tabular format and exported in ASCII format to common spreadsheet applications, minimizing clinical data preparation.

Efficient Beam Model Validation

To validate beam models, the physicist can calculate and compare depth dose curves and profiles in a water phantom corresponding to input data. Eclipse displays calculated data superimposed over the measured data, eliminating the time-consuming task of printing and overlaying plots. Percentage difference and distance-to-agreement values are displayed as recommended by the AAPM Task Group 53 on Quality Assurance for Clinical Radiotherapy Treatment Planning.

Built-in safeguards prevent beam models from being used for dose calculations until approved by an authorized user. Eclipse performs cyclical redundancy checks (CRCs) on all beam data files each time they are accessed. If a failure is detected, an error message is issued and further dose calculation is prevented until the condition is corrected and the beam data is re-approved.

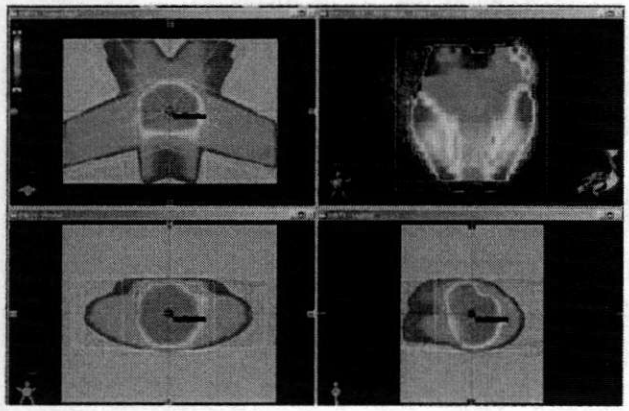


Eclipse displays percentage difference and distance-to-agreement values to compare calculated and measured beam data.

Eclipse™ Treatment Planning System

Integrated Plan Verification and Quality Assurance

A critical part of pre-treatment quality assurance is the verification of treatment plans and their corresponding dose distributions. In Eclipse, treatment plans can be copied and calculated on any geometrical or scanned phantom. This process is automated and accelerates quality assurance tasks.

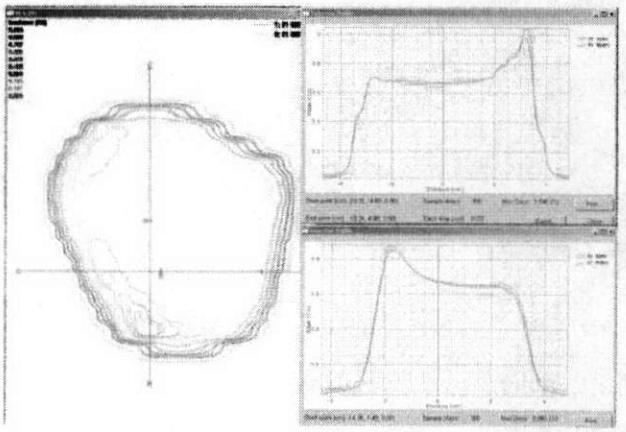


A five-field IMRT plan is calculated on a water phantom for pre-treatment verification.

Dose calculated in the phantom is compared with point dose or dose distribution measurements. Dose distributions in any plane through the phantom can be displayed and exported to third-party dosimetry QA systems via DICOM RT. Exported objects include plan geometry, dose prescription, reference points, blocks or MLC apertures, wedge information, intensity-modulated radiation therapy (IMRT) fluence, and 2D and 3D dose distributions.

Eclipse can calculate a calibrated dose distribution at the portal imager. The physicist can compare this to an electronic dose image from Varian's PortalVision™ MV imaging system, eliminating phantom setup and film processing and offering an integrated quality assurance method.

Eclipse is part of the Inspiration™ integrated oncology environment, where all Varian products share a common database. Verification plan data is automatically stored with the patient record, ensuring that the correct patient-specific QA data is used. The Inspiration environment eliminates the additional quality assurance required when exporting data to non-integrated systems.



Calculated and measured portal dose images are analyzed using line dose profiles.

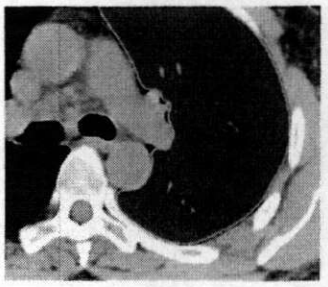
VARIAN medical systems | ONCOLOGY SYSTEMS

USA Headquarters, California
Varian Medical Systems
Palo Alto, CA
Tel: 650.424.5700
800.544.4636
Fax: 650.424.5823
www.varian.com

Headquarters Europe, Eastern Europe, Africa, Middle & Near East
Varian Medical Systems International AG
Zug, Switzerland
Tel: 41.41.749.8844
Fax: 41.41.749.8844
info.europe@varian.com

RAD7151B

7/06 (SK)



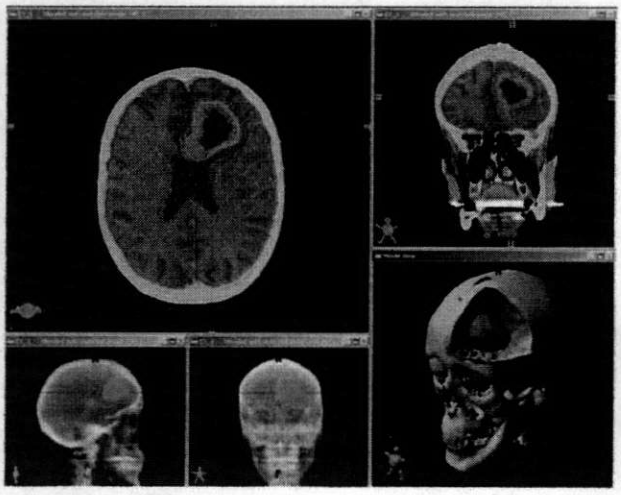
Eclipse™ Treatment Planning System

Contouring

For intensity-modulated radiation therapy (IMRT) and 3D conformal therapy planning, segmenting organs at risk and accurately delineating target volumes are critical. Structure segmentation time is reduced from hours to minutes with the powerful contouring tools in Eclipse. Clinicians can accurately define target and critical structures on fused multi-modality images with advanced drawing and editing capabilities. Enhanced templates and powerful post-processing of structures accelerate the contouring process.

Comprehensive Image Visualization

Targets and structures can be accurately defined using multiple image modalities, such as CT, MR, cone-beam CT, PET, and SPECT. Images can be fused or registered manually using fiducial markers or anatomical landmarks. Either DICOM coordinates or mutual information may be used for automatic matching. The clinician can interactively adjust the relative weights and colors of the different images to better visualize soft tissue or functional information while contouring. Structures and targets defined on one image set may be copied and pasted onto another image set, accelerating structure segmentation especially for 4D CT images.



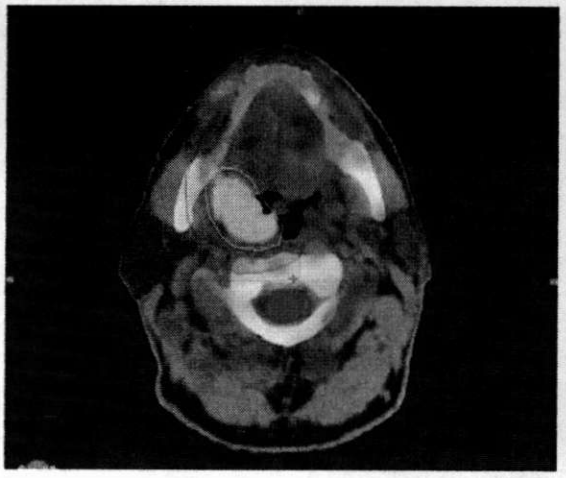
Fused images can be blended to better visualize tumors or critical structures during contouring.

Eclipse renders segmented volume data with user-specified colors and styles to enhance visual interpretation of 3D image data. In addition to the traditional transverse CT view, clinicians can display and work on coronal or sagittal views. A non-divergent digitally reconstructed radiograph (DRR) simplifies navigation through a series of CT images. Using model views, the clinician has a room's eye view of all patient structures rendered in 3D for easy spatial orientation. Image data is easily interpreted in Eclipse using window leveling and image processing, which is crucial for accurately defining structures of concern.

Flexible Structure Definition

In the contouring process, defining structure names and properties is tedious and repetitive. Structure templates in Eclipse make this process more efficient. These templates are based on an anatomical database that includes the structure name, display properties, ICD code, volume type, and automated segmentation parameters. Clinicians can create custom templates and modify existing ones. Structure templates are part of the clinical protocol templates that accelerate the entire planning process by pre-populating all of the plan parameters from physician's intent. These customizable protocol templates promote standardized clinical protocols by ensuring the same structures are contoured in the same way from patient to patient for any treatment site.

Eclipse completely supports the DICOM RT Structure Set object. Therefore, structure sets imported from other virtual simulation and treatment planning systems are reproduced accurately. Patient structures can be shared or transferred to other systems through DICOM RT. To facilitate DART™ dynamic adaptive radiation therapy and image-guided radiation therapy (IGRT), the user can copy structures from one image set to another registered image set.



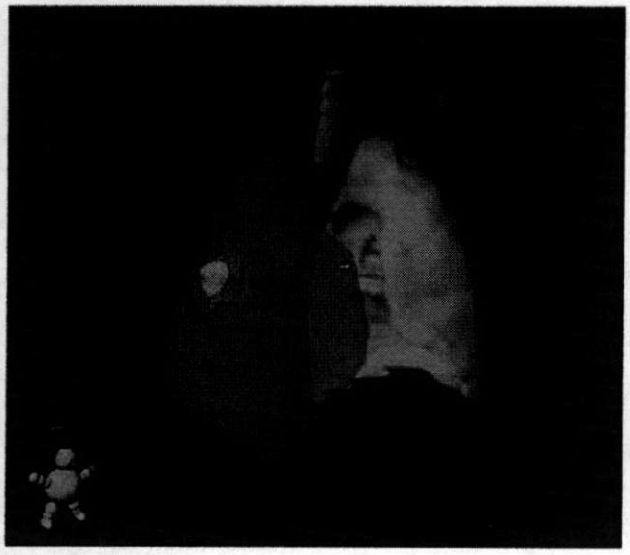
Fused PET and CT images highlight a head and neck tumor.

Eclipse™ Treatment Planning System

Efficient Segmentation

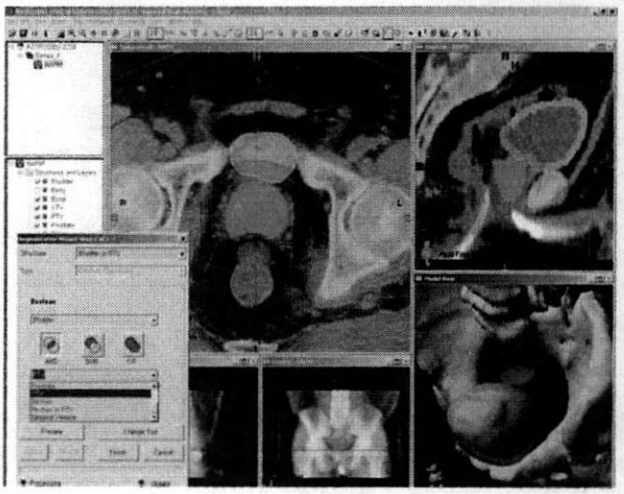
Defining targets and organs at risk is a large part of the segmentation process. This is accelerated by supporting the clinician's preferences — from drawing freehand with the all-in-one mouse tool to drawing volumetrically using an adaptive paintbrush that is sensitive to differences in image values and gradients. The user can easily zoom, pan, and scroll through image data using the mouse wheel.

Eclipse uses both logic-based and image-based automatic segmentation to increase speed and efficiency. Logic-based segmentation automatically creates new structures by forming Boolean combinations of existing structures and extracting walls or shells of a given thickness. For example, automatic cropping of structures to fit within the body saves time when correcting a PTV expansion. Image-based segmentation automatically fills regions based on organ-specific parameters or a user-defined range of CT values. Segmentation from one CT slice can be automatically extended and adapted to the next slice. A segmentation wizard contours some structures with a single click of the mouse.



Using the segmentation wizard, the lungs and spinal cord are contoured automatically.

The clinician can confine the action of automated segmentation tools to a rectangular or free-form volume of interest. For example, applying automatic segmentation based on a CT value within the user-defined volume of interest quickly completes the typically time-consuming task of contouring the femoral heads.



Using Boolean logic, dosimetrists create custom optimization structures for a prostate IMRT treatment.

Powerful Editing and Post-Processing

Much of the extensive labor associated with editing, modifying and "cleaning up" segmentation work is eliminated with the advanced editing and post-processing capabilities of Eclipse. In addition to freehand outline correction and cut/copy/paste contour editing, Eclipse has automatic clean-up, extraction, and enhancement functions to process structures after they have been defined.

These functions remove, keep, connect, or disconnect pieces of structures either by manual selection or automatically based on size. For example, the treatment couch and head-frame in images of a head and neck treatment case can be eliminated from every CT slice in just one processing step. Structures in Eclipse can also be trimmed at the boundary of other structures, smoothed, and cavities filled according to criteria provided by the clinician.

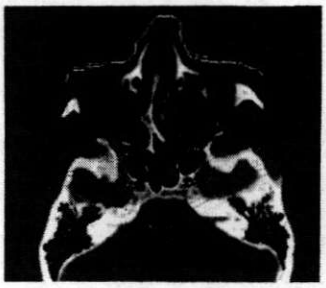
VARIAN medical systems | ONCOLOGY SYSTEMS

USA Headquarters, California
Varian Medical Systems
Palo Alto, CA
Tel: 650.424.5700
800.544.4636
Fax: 650.493-[illegible]
www.varian.com

Headquarters Europe, Eastern Europe, Africa, Middle & Near East
Varian Medical Systems International AG
Zug, Switzerland
Tel: 41.41.749.8844
Fax: 41.41.749-[illegible]
info.europe@varian.com

RAD7153B

7/06 (5K)



Eclipse™ Treatment Planning System

Dose Calculation

Advanced algorithms in Eclipse quickly and accurately calculate dose distributions for photons, electrons, and protons. With the combination of modular algorithms and the flexible architecture of Eclipse, clinicians can select the optimum algorithm for each treatment modality. The clinician can rapidly customize intensity-modulated radiation therapy (IMRT) plans using the Dose Volume Optimizer (DVO). A flexible calculation framework increases efficiency in the treatment planning process particularly in a distributed planning environment. Specialized beam data analysis tools reduce the time for commissioning.

Photons

Eclipse has a variety of options for photon dose calculation, including simple 2D point dose, advanced 3D convolution superposition, inverse-planned IMRT optimization and portal dose calculation. Long known for its fast and accurate calculations, the Pencil Beam Convolution (PBC) algorithm continues to support all photon beam treatment modes and has several methods for heterogeneity correction.

The Anisotropic Analytical Algorithm (AAA) is a Monte Carlo-based convolution superposition algorithm that covers the entire therapy range. AAA accurately models dose deposition in regions with a high degree of tissue heterogeneity by accounting for the 3D density variations directly in the dose calculation.

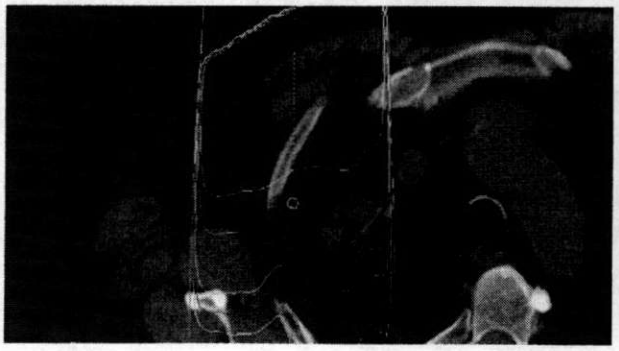
With Eclipse's IRREG calculation, simple 2D plans can be quickly created with or without images. Dose is prescribed to a specified reference point and the corresponding MU is calculated in addition to doses at other points of interest. For maximum flexibility, IRREG supports wedges, blocks, and multileaf collimators (MLCs). It is available as an option on the Varian Acuity™ treatment planning, simulation and verification system.

The Portal Dose Calculation (PDC) algorithm calculates a dosimetric portal image that can be compared with the dose measured by Varian's PortalVision™ MV imaging system. The PDC achieves very accurate dose images by using a kernel developed specifically for the amorphous silicon detector. Within Varian's Inspiration™ integrated oncology environment, the PDC and PortalVision simplify pretreatment verification of complex IMRT plans by eliminating the time-consuming process of film dosimetry.

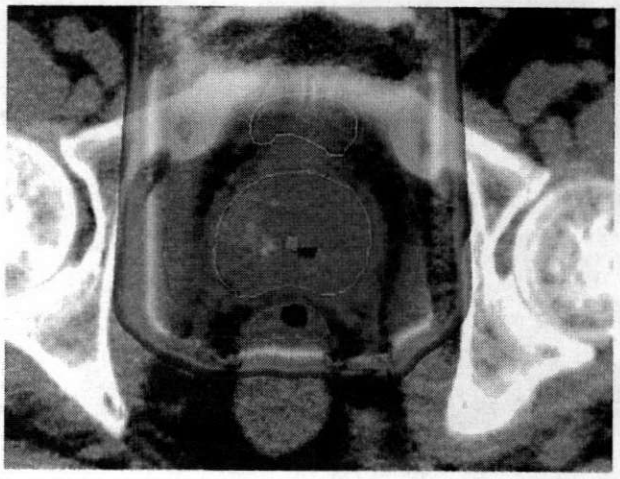
Electrons

Eclipse uses the Generalized Gaussian Pencil Beam (GGPB) algorithm or the optional electron Monte Carlo (eMC) algorithm for calculating electron dose distributions. Both algorithms support irregular field shapes, collimator rotations and non-coplanar fields.

The eMC algorithm is based on standard EGSNRC Monte Carlo methods and achieves comparable accuracies in one-tenth the time. By modeling heterogeneous tissue with pre-simulated spheres of various sizes and densities, eMC decreases the number of electron transport steps through the patient. This reduces the calculation time without sacrificing accuracy even in very heterogeneous regions.



AAA calculated single field isodose distribution demonstrates heterogeneity corrections.



View shows a dose distribution of a proton beam to treat prostate cancer.

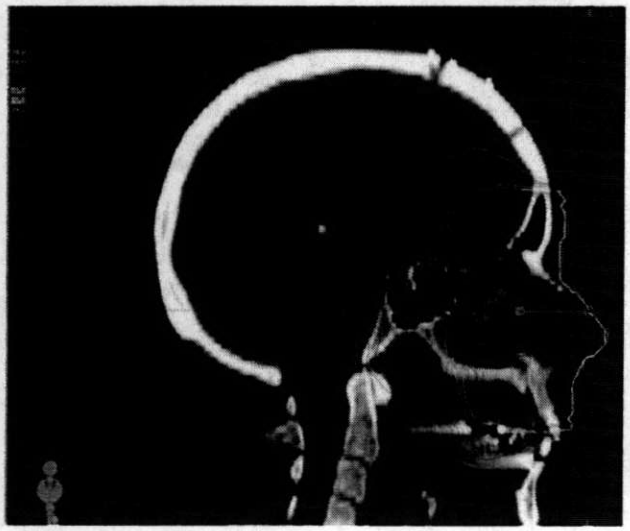
Eclipse™ Treatment Planning System

Protons

Eclipse Proton Planning uses a unique 3D pencil beam convolution superposition algorithm for fast and accurate dose calculations. Dose kernels derived from Monte Carlo calculations within the patient are convolved with in-air proton fluences. Consequently, this algorithm is easily adapted for both double-scattering and wobbling beam lines and supports proton accelerators from a variety of manufacturers. The 3D Analytical Pencil Beam algorithm accurately models heterogeneities, which are very important for proton calculations.

Rapid Optimization

The DVO creates highly conformal dose distributions by optimizing beam intensity profiles according to user-defined dose-volume objectives and the Normal Tissue Constraint. The DVO uses simple gradient optimization with line minimization to iteratively modify the field fluences and recalculate the dose distribution until the desired objectives are reached.



Electron Monte Carlo dose distribution is displayed on a sagittal view.

During each iteration, Eclipse uses the Multi-Resolution Dose Calculation (MRDC) algorithm for a 3D dose calculation that accounts for full scatter and tissue heterogeneities. The MRDC is based on the 3D convolution of Monte-Carlo-generated point-spread function kernels. Because of the high speed of the MRDC algorithm (a few hundred milliseconds to calculate an IMRT field), the clinician can interact with the optimization in real-time, rapidly customizing the plan for each patient.

After dose optimization, Eclipse calculates the leaf motions needed to deliver the dose defined by the field fluences. In this calculation, Eclipse accounts for the physical properties of the MLC and maximizes delivery efficiency while minimizing deviations from the optimal fluence. With the resulting MLC leaf sequence, the final dose calculation accurately represents the dose delivered by the accelerator.

Flexible Calculation Framework

Eclipse uses a flexible dose calculation framework to manage calculation activities across a distributed network. The clinician can designate certain workstations within the Eclipse network for local or remote calculations. The Eclipse Calculation Queue computes dose distributions in the background and stores the results for later retrieval, thereby freeing the clinician to work on other tasks.

With multiple algorithms to choose from, the clinician can select the one best suited for each clinical case. When new versions of algorithms are released, the physicist can compare versions side-by-side for commissioning. New algorithms can be implemented as they become available without having to upgrade the entire treatment planning system. Varian has developed an applications programming interface (API) so researchers can implement their own algorithms within Eclipse.

Easy Commissioning

Eclipse is easy to commission. Configuring photon beams from basic input data is completely automated for AAA. All Eclipse algorithms have minimal beam data measurement requirements. This reduces the labor of collecting basic input data and configuring new beam models. The physicist can begin using new algorithms quickly and confidently, because the Dose Calculation Framework supports side-by-side algorithm calculation comparisons.

Pre-configured AAA beam models and pre-measured basic input data for PBC and GGPB are available for the Varian Clinac® EX (and most C/CD) high-energy linear accelerators. This significantly reduces beam model configuration time. Beam data analysis tools in Eclipse simplify the validation of beam models and calculations. Eclipse automatically calculates depth dose curves and profiles in a water phantom corresponding to each measured input data curve. Comparison data are presented as recommended by AAPM Task Group 53 on Quality Assurance for Clinical Radiotherapy Treatment Planning.

VARIAN medical systems | ONCOLOGY SYSTEMS

USA Headquarters, California
Varian Medical Systems
Palo Alto, CA
Tel: 650.424.5700
800.544.4636
Fax: 650.493.5634
www.varian.com

Headquarters Europe, Eastern Europe, Africa, Middle & Near East
Varian Medical Systems International AG
Zug, Switzerland
Tel: 41.41.749.8844
Fax: 41.41.740.5340
info.europe@varian.com

RAD7161B

7/06 (5K)

2/16

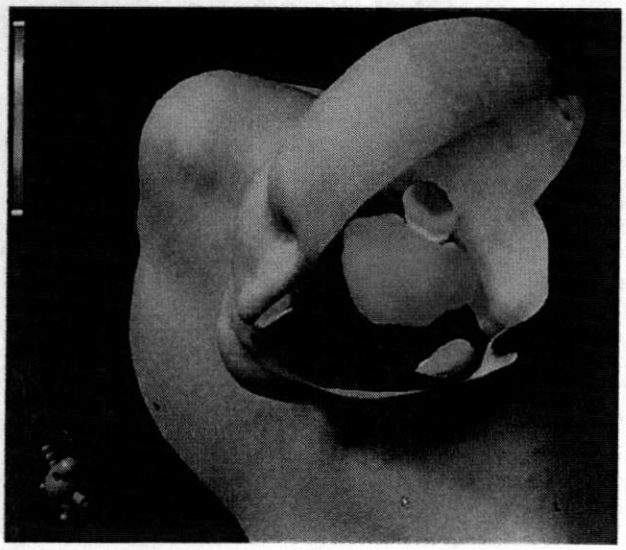
Eclipse™ Treatment Planning System

Plan Evaluation

Comparing different plan modalities helps an oncology team decide on the optimal method for therapy. With Eclipse, clinicians can combine, compare, and evaluate different treatment modalities including 3D conformal, intensity-modulated radiation therapy (IMRT), electron, proton, and brachytherapy on a single integrated planning system. They can also conveniently compare candidate plans in order to decide on the most effective course of patient treatment. Eclipse workstations can be distributed throughout a clinic so plans can be evaluated anywhere, anytime, which speeds up the planning process.

Multiple Dose Displays

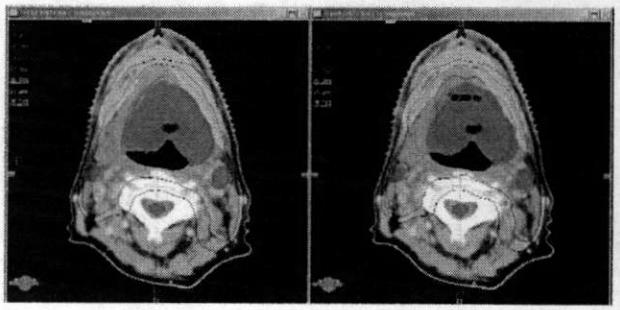
Clinicians can evaluate plans on fused multi-modality images with customizable 2D and 3D dose displays. In the beam's eye view (BEV) and 3D views, Eclipse renders isodose surfaces as wire frames or solid surfaces with user-selected colors, transparencies, and values. Interactive field-weight adjustments automatically update all dose displays giving the clinician immediate feedback. Dose cloud and surface dose renderings complement differential and cumulative dose-volume histograms (DVHs), as an effective way of visualizing target coverage and the dose to critical structures. By using the show crosshair tool with the DVH display, the clinician can point at any part of a DVH curve and the isodose level will display on the orthogonal plan views. This highlights the spatial distribution of dose in conjunction with the DVH display to facilitate the physician's review of the plan. The clinician can choose DVH line styles, colors, and thicknesses to better visualize DVH calculations.



Surface dose rendering of a head and neck IMRT plan displays the dose distribution on the surface of the tumor, adjacent brain stem, and left parotid gland.

Powerful Plan Comparison

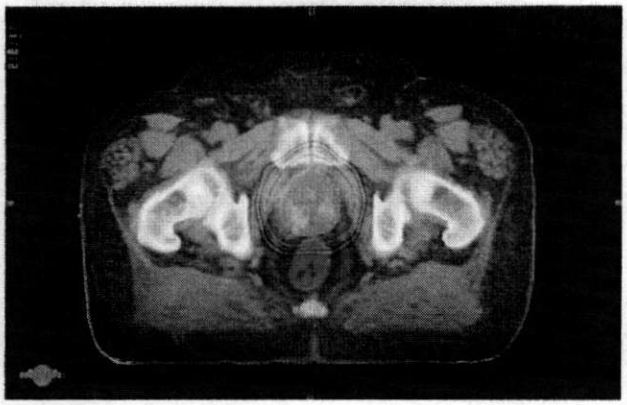
With flexible screen layouts and intuitive plan comparison tools, the clinician can determine the best course of treatment for every patient. For fast and simple plan comparisons, side-by-side plan displays can be locked together to ensure that the same view from each plan is used during evaluation.



Axial views of a head and neck plan compare the dose distributions of IMRT sliding window and step-and-shoot delivery.

In addition to the qualitative side-by-side dose comparisons, DVHs for up to six plans can be displayed on the same graph for quantitative analysis. The clinician can compare the axial dose distributions for these six plans on the same screen. Using Boolean operators the clinician can calculate the DVHs for combinations of structures. The differences in two dose distributions can be quickly evaluated using plan subtraction. With plan summation, the clinician can evaluate the total dose distribution of a multi-phase course of treatment. These tools can be used across multiple treatment modalities to evaluate, for example, combined brachytherapy and external beam plans, or multiple courses of treatment.

Eclipse™ Treatment Planning System



Cone-beam CT images can be fused and blended with the original CT images and used for DART.

For the ultimate treatment record and for DART™ dynamic adaptive radiation therapy, Eclipse can calculate the dose that was delivered to the patient using the actual multileaf collimator (MLC) positions and cone-beam CT (CBCT) even for a partial treatment. The dose distributions and DVHs can be compared to assess the evolution of the treatment over time. In this way, the clinician makes decisions about adapting the treatment strategy to the changing patient anatomy. If a change is needed, the plan can be quickly recreated for the CBCT image and used for subsequent treatments.

Eclipse tabulates information about dose prescription and fractionation, and summarizes dose contributions to reference points. Consequently, the clinician has an overview of the whole treatment strategy.

Eclipse displays the combined dose distributions of a four-field cervix plan with a brachytherapy boost.

Flexible Plan Documentation

Eclipse generates thorough plan documentation that meets the clinic's unique requirements for quality assurance. Hard-copy reports include BEV plots, isodose plots, and screen printouts. Clinicians can customize plan reports using a simple text editor, and they can incorporate screen captures into presentations and technical papers. DVHs can be plotted, printed in tabular form, or exported in ASCII format suitable for external analysis software.

Electronic Plan Approval

After selecting the ideal treatment plan, the oncologist approves the plan electronically. The approval is password-protected. An approved plan cannot be unintentionally modified, ensuring its integrity. When Eclipse is part of Varian's Inspiration™ integrated oncology environment, the approved plan becomes immediately available for treatment. Safety is enhanced by eliminating data transfers from treatment planning to treatment delivery. Plans can be easily converted into clinical protocol templates for reuse as class solutions for future patients.

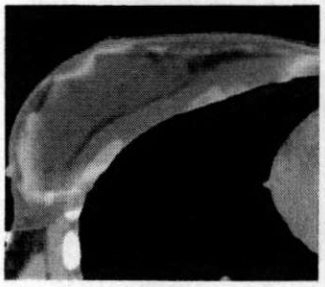
VARIAN medical systems | ONCOLOGY SYSTEMS

USA Headquarters, California
Varian Medical Systems
Palo Alto, CA
Tel: 650.424.5700
800.544.4636
Fax: 650.493.5627
www.varian.com

Headquarters Europe, Eastern Europe, Africa, Middle & Near East
Varian Medical Systems International AG
Zug, Switzerland
Tel: 41.41.749.8844
Fax: 41.41.749.5840
info.europe@varian.com

RAD7159B

7/06 (5K)



Eclipse™ Treatment Planning System

Eclipse Interactive IMRT Planning

Powerful 3D conformal planning tools in Eclipse are combined with interactive dose-volume optimization for fast, flexible, and accurate intensity-modulated radiation therapy (IMRT) planning. Clinicians can make real-time decisions using the interactive IMRT features in Eclipse. With thousands of patients at hundreds of institutions worldwide successfully treated with IMRT plans generated by Eclipse, it is an established solution for IMRT planning.

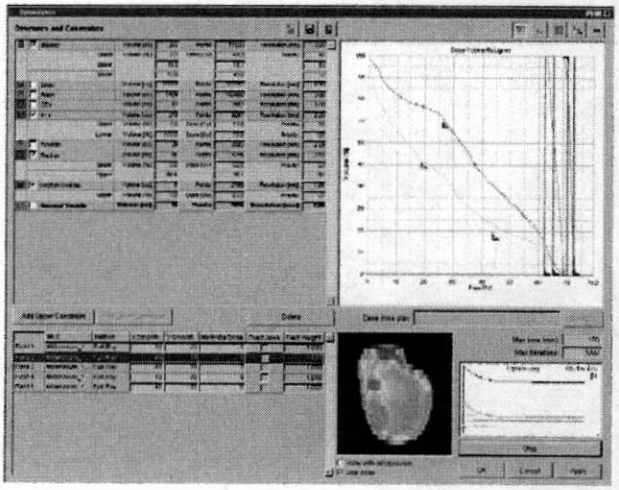
Rapid IMRT Plan Setup

Eclipse has comprehensive clinical protocol templates that speed the planning process. These templates pre-populate all of the planning parameters based upon the physician's intent, including structures, field arrangements, optimization criteria, and even output formatting. With clinical protocol templates, clinicians can create class solutions for IMRT that can be easily individualized for the needs of each patient through interactive optimization. Easy transfer between institutions simplifies the sharing of clinical experience and the adoption of IMRT.

The complexity of IMRT demands the segmentation of many organs at risk and accurate delineation of target volumes. Powerful tools in Eclipse such as enhanced structure templates, advanced contouring tools, automatic segmentation, structure post-processing, and flexible screen layouts, reduce segmentation time from hours to minutes.

Clinical Flexibility

With the flexibility of Eclipse, clinicians can rapidly customize treatment plans for individual patients. Eclipse can incorporate the base dose from an existing treatment plan to optimize IMRT boosts or to adapt the plan based upon the treatment already delivered. By accounting for these base dose distributions, the clinician can combine IMRT with conventional photon, electron, proton, or brachytherapy plans. Eclipse automatically splits the fields for large field IMRT. For lateral fields in head and neck treatments, the jaws can be locked to decrease the dose given to shoulders. Planning for non-coplanar beams and multiple targets simplifies even the most complex cases.



The user interface updates optimization information with each iteration so the clinician can interactively modify parameters based on real-time feedback.

Interactive Optimization

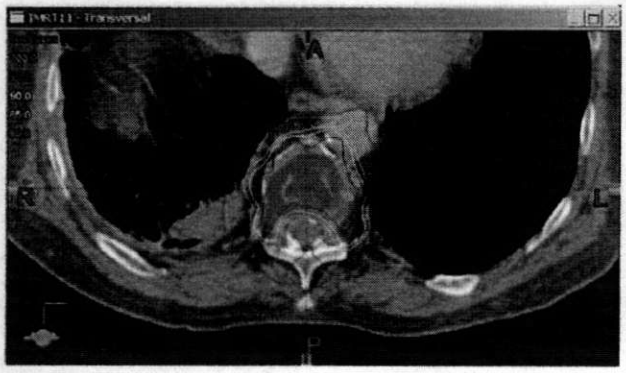
Eclipse uses a very fast optimization algorithm to converge rapidly to a good plan. While the plan evolves, the clinician can see real-time updates to the dose-volume histogram (DVH), objective function, and fluence matrices. The best plan can be created quickly by interactively modifying the dose constraints during optimization. The first plan is the right one every time with the Normal Tissue Constraint, which removes the need for "tuning" structures. The clinician can restart any previous optimization from where it ended, even using different image sets, to efficiently adapt treatment plans during the course of treatment.

Eclipse™ Treatment Planning System

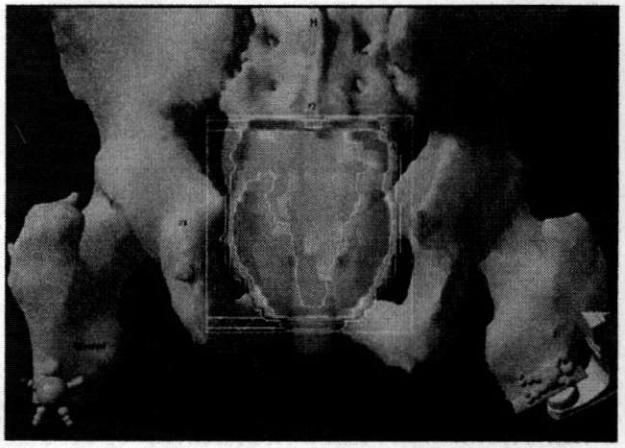
Accurate Dose Calculation

During each iteration of the optimization, Eclipse's fast algorithm accurately calculates dose to the entire volume, including bolus, while accounting fully for scatter and heterogeneities. DVH curves displayed during optimization match the results for the final dose calculation.

The leaf motion calculator (LMC) converts optimal fluence to a deliverable leaf motion sequence and fluence while reducing treatment time and monitor units. The LMC incorporates the physical limitations, leakage, and dosimetric effects of the multileaf collimator into the leaf sequence calculation.



Eclipse IMRT plan shows good sparing around the spinal cord.



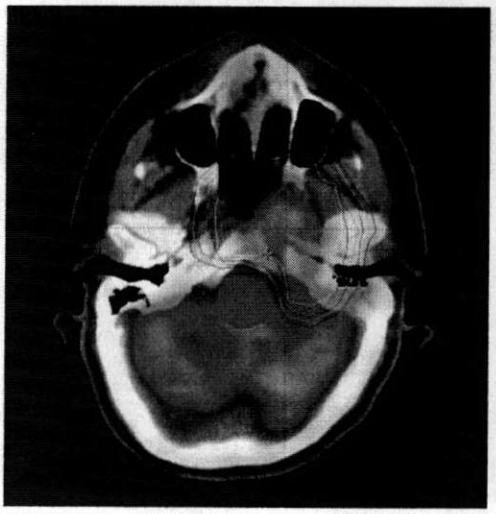
Fluence matrix displayed in the beam's eye view (BEV) highlights the correlation with the underlying patient anatomy.

Delivery Efficiency

Sliding window IMRT deliveries for Varian C-Series Clinac® linear accelerators and multiple static segment treatments for Varian, Siemens, Elekta, Mitsubishi, and General Electric accelerators can be calculated in Eclipse. High-resolution IMRT treatments planned on Eclipse can be delivered in a normal 15-minute time slot on a Varian Clinac.

Efficient Plan Verification and QA

Eclipse can calculate QA dose distributions for pre-treatment plan verification on either QA phantoms or the PortalVision™ MV imaging system. Comparing the resulting dose distributions with point dose measurements, film dosimetry, and PortalVision dosimetric images, greatly reduces the QA time required for IMRT pre-treatment verification. With extensive support of the DICOM RT standard, the clinician can export all pertinent plan information from Eclipse to independent monitor unit calculation programs or verification devices, and for clinical protocol studies.



IMRT dose distribution on a fused CT/PET image shows brain stem sparing.

VARIAN medical systems | ONCOLOGY SYSTEMS

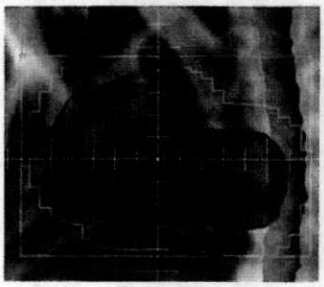
USA Headquarters, California
Varian Medical Systems
Palo Alto, CA
Tel: 650.424.5700
800.544.4636
Fax: 650.493.5627
www.varian.com

Headquarters Europe, Eastern Europe, Africa, Middle & Near East
Varian Medical Systems International AG
Zug, Switzerland
Tel: 41.41.749.8844
Fax: 41.41.7405340
info.europe@varian.com

RAD7152B

7/06 (5K)

2520



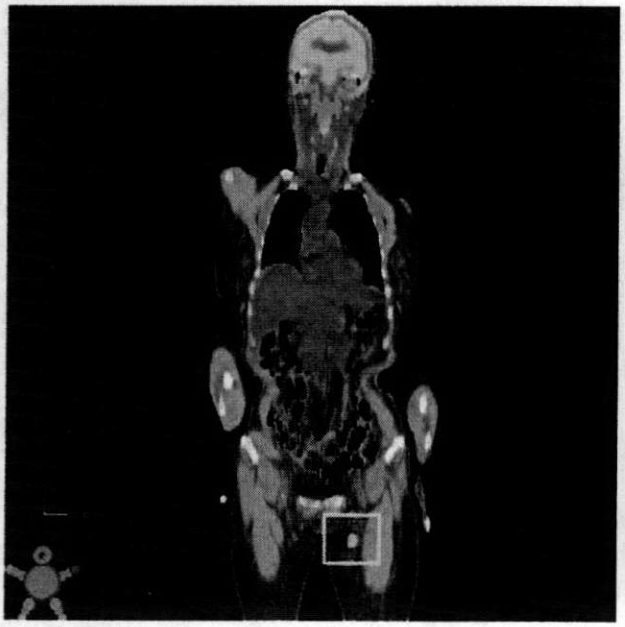
Eclipse™ Treatment Planning System

Virtual Simulation and Field Setup

With Eclipse, virtual simulation and treatment planning are integrated in one comprehensive system. Virtual simulation and planning processes are accelerated through excellent segmentation software with clinical protocol templates, hot-keys, automatic field positioning, automatic field aperture shaping, and a flexible interactive user interface. Efficiency is improved throughout the simulation and planning processes when Eclipse is integrated with the Acuity™ treatment planning, simulation and verification system, the PortalVision™ MV imaging system, and the On-Board Imager® kV imaging system (OBI) from Varian.

Versatile Image Management

Accurate patient modeling for treatment planning is critical. Eclipse can create 3D patient models from any DICOM 3.0-compliant image set, including CT, MR, and PET. Using image fusion, the clinician can precisely define targets and critical structures.

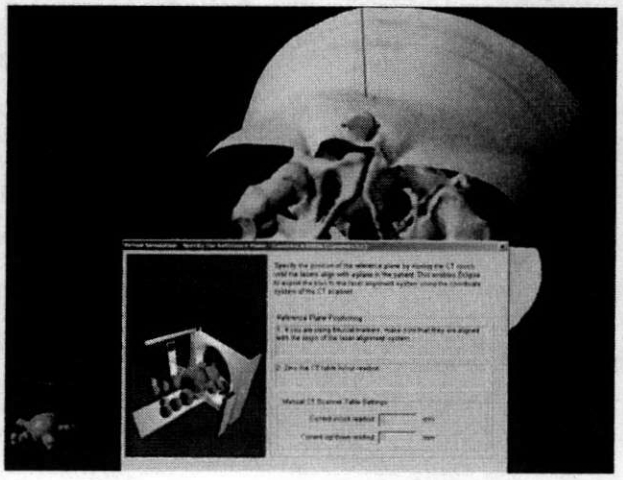


Inguinal nodes are targeted more precisely using fused CT/PET images.

To verify patient positioning, digitally reconstructed radiographs (DRRs) are generated in Eclipse and compared with setup images. To adapt to clinical changes in patient anatomy and position on a day-to-day basis, the clinician can adjust the patient's treatment plan based on cone-beam CT images taken on the accelerator. To account for organ motion during respiration, 4D and maximum intensity projection (MIP) images can be imported into Eclipse.

Comprehensive Virtual Simulation

With Eclipse installed in a CT simulation room, clinicians can do real-time or off-line simulation with or without contouring. Isocenter and field parameters are transferred to multiple laser systems such as LAP and Gammex for easy patient marking. The isocenter position can be located relative to the simulation isocenter with user-definable coordinates. Virtual simulation information is saved in the Eclipse database and can be accessed from any Eclipse workstation located in the department for other planning tasks.

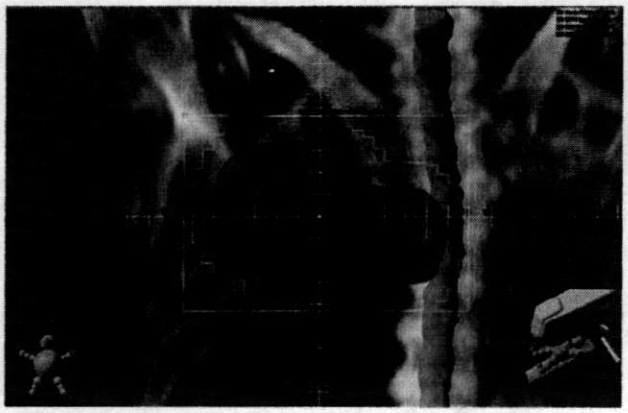


The isocenter and field entrance positions are exported to the moveable laser system for marking the patient.

The contouring step of the virtual simulation process is easy and efficient with Eclipse. A wide range of automatic, semi-automatic, and manual tools simplify the segmentation task and match any drawing preference. For example, a segmentation wizard contours some structures with a single click of the mouse. Intelligence is built into even the simplest tools so the oncologist spends less time meticulously outlining each organ of interest. For different perspectives, the oncologist can pan and zoom with the mouse and contour in coronal, sagittal, and axial planes.

Eclipse™ Treatment Planning System

Generating DRRs is an important aspect of virtual simulation. Eclipse updates the DRR image in real-time whenever the beam geometry is changed. From a library of customizable parameters, multi-channel DRRs highlight airways, soft tissue, and bony structures in the same image. To support virtual fluoroscopy, Eclipse calculates real-time DRRs with structure, graticule, field aperture overlays, and graphical field design functions.

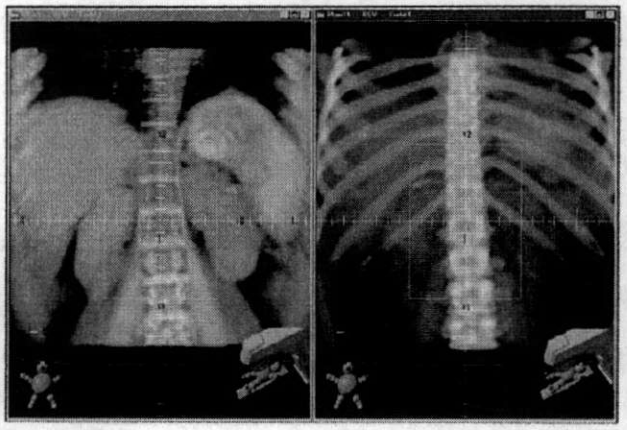


Virtual fluoroscopy is easy using volume rendering of structures in conjunction with DRRs in a BEV.

Efficient Field Setup

Clinicians can set up fields during the simulation process or at a later time anywhere in the department. The clinician can define fields graphically in planar, beam's eye (BEV), or 3D model views. The isocenter can be relocated, the beam geometry adjusted, and field apertures defined all with the mouse. Field parameters are displayed in a spreadsheet-like manner for easy entry or editing. Hot-keys and single-click functions simplify multi-step tasks. After the field arrangement is modified, clicking a single button reshapes all field apertures to fit to the target volume.

Eclipse clinical protocol templates increase planning efficiency and clinic-wide standardization of treatment techniques by tying together everything from physician's intent to plan review preferences in a single location. For virtual simulation, these protocol templates apply prescription information, define structure sets, center treatment fields, conform field apertures, and set calculation model options.



Two images of a thorax patient are compared. On the left, soft tissue is enhanced while bony anatomy is suppressed, and on the right, bony anatomy is enhanced.

The clinician can modify stored templates. Any existing plan can be converted into a clinical protocol template and conveniently organized into custom folders ready to be applied to subsequent patients.

Unmatched Data and Process Integration

Virtual simulation and treatment planning are integrated in Eclipse. For example, emergency plans can be created and doses calculated in the simulation room, saving precious time. Data integration eliminates the need for a separate virtual simulation database. All patient information is stored in one central database, eliminating transfer of data from one application to another. DRRs generated in Eclipse are readily available as reference images for pre-treatment verification on Acuity and for daily imaging using PortalVision and OBI.

A treatment plan can be electronically approved in Eclipse, preventing modification and ensuring plan integrity. When Eclipse is part of the Varian Inspiration™ integrated oncology environment, the approved plan immediately becomes available for treatment. With extensive support of the DICOM standard, treatment plan information can be accessed by multiple vendor devices.

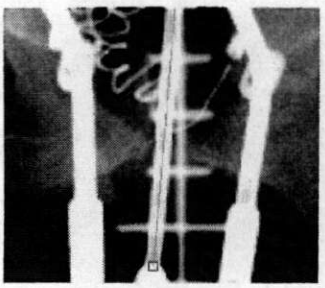
VARIAN medical systems | ONCOLOGY SYSTEMS

USA Headquarters, California
Varian Medical Systems
Palo Alto, CA
Tel: 650.424.5700
800.544.4636
Fax: 650.493.1053
www.varian.com

Headquarters Europe, Eastern Europe, Africa, Middle & Near East
Varian Medical Systems International AG
Zug, Switzerland
Tel: 41.41.749.8844
Fax: 41.41.758.3510
info.europe@varian.com

RAD7156B

7/06 (5K)



Eclipse™ Treatment Planning System

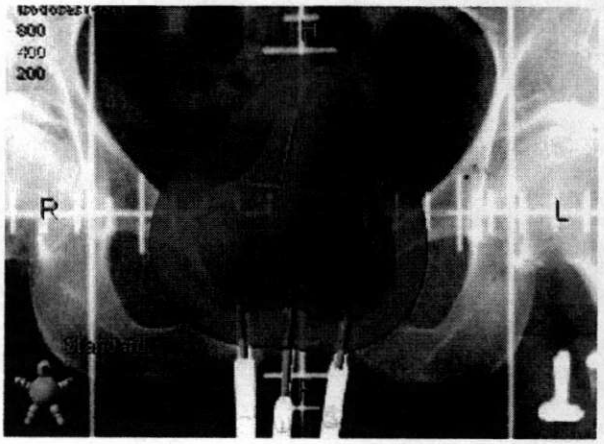
Brachytherapy

The combination of the BrachyVision™ 3D™ treatment planning system with the VariSource™ 200 HDR afterloader and GammaMed*plus*™ HDR/PDR afterloader establishes Varian as a leader in brachytherapy. In addition to being sold as a standalone brachytherapy planning system, BrachyVision can also be part of an Eclipse treatment planning system. For those customers who wish to perform film-based brachytherapy planning, Eclipse includes BrachyVision 2D. BrachyVision 3D is available as an option for Eclipse users who require image-guided 3D brachytherapy planning. Information on BrachyVision 3D can be found in a separate brochure.

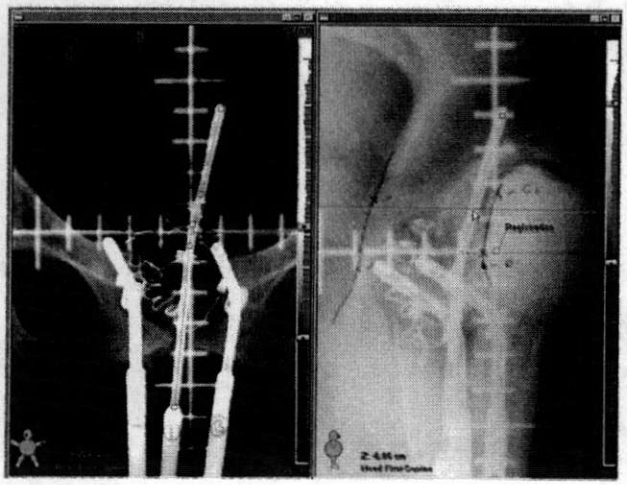
Film-based Brachytherapy Planning

BrachyVision 2D in Eclipse accepts plan data entered using digitizer tablets, film scanners, or other 2D image sources. Reference images can be orthogonal or semi-orthogonal pairs, isocentric shift films, or films created using custom reconstruction jigs. It is also possible to enter plan data manually using coordinates.

Multiple film sets are supported for more complex plans. If scanned films are used for planning, BrachyVision 2D displays the location of sources projected onto the film planes. This improves source positioning accuracy and aids in rapid identification of individual sources. As part of Varian's ARIA™ oncology information system, BrachyVision 2D gives Eclipse clinicians the advantage of direct use of fluoroscopic (or flat panel) images from the Acuity™ treatment planning, simulation and verification system – greatly speeding and simplifying the planning process.



Dose is displayed over film for aid in visualization.

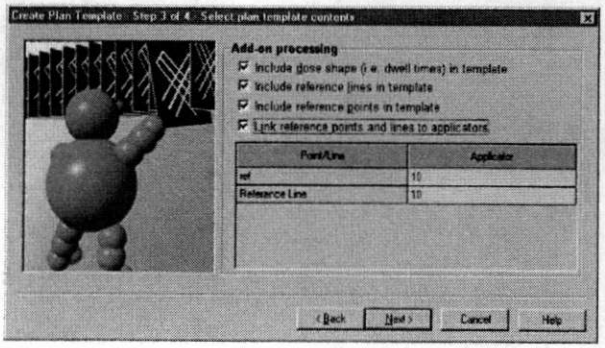


Digitized film images for aid in data entry.

Supported Sources and Techniques

BrachyVision 2D supports the use of any brachytherapy source for which TG43 format data is available, whether as individual sources or used in afterloading devices. HDR, MDR, LDR, and PDR treatment modalities are all supported. Clinicians can plan according to Manchester or Paris rules, with automatic calculation of and normalization to basal dose points.

Preplanning may be carried out using implant templates or plan templates. The clinician can experiment with a number of configurations in an imageless plan before implanting the patient.

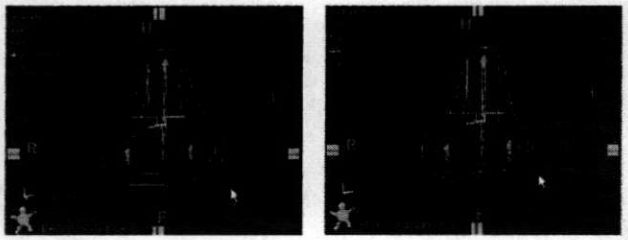


BrachyVision has the ability to store template plans.

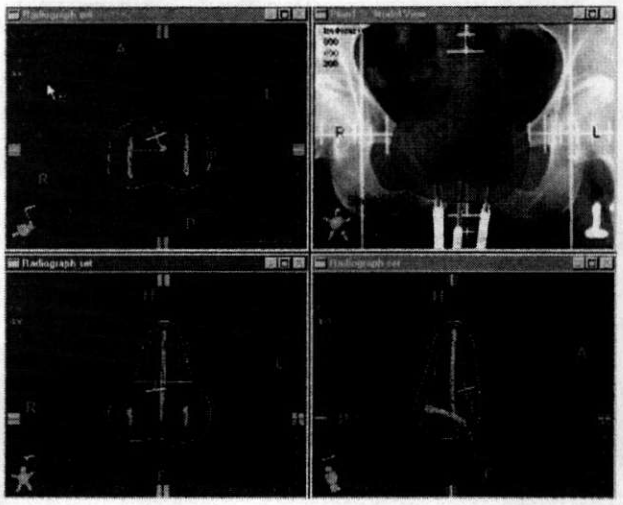
Eclipse™ Treatment Planning System

Flexible Plan Normalization

To achieve precise control over the delivered dose, BrachyVision provides several options for adjusting the dose distribution, such as normalization to a user-defined point, dose rescaling, and dose shaping. Interactive renormalization can be carried out using the Dose Rescaler™ tool, which allows the clinician to pull the isodose distribution to the desired size. For plan types where dwell times change within a single applicator (e.g., HDR with a stepping source), the planner can shape the dose distribution simply by pushing or pulling isodose lines using the unique Dose Shaper™ tool.



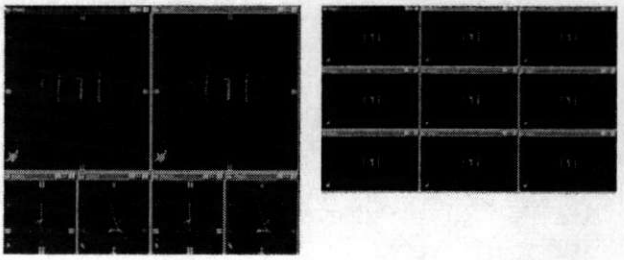
The Dose Shaper tool provides interactive dose optimization.



3D over film view helps to visualize anatomy.

Plan Analysis

Dose distributions in Eclipse are easy to evaluate using the extensive range of tools available. For fast and simple plan comparisons, any two plans can be compared side-by-side during plan evaluation. The clinician can also compare and combine plans of different modalities using Eclipse (e.g., brachytherapy and external beam plans). It is possible to manipulate the 2D orthogonal planes and the 3D dose model with zoom, rotate, and pan features in order to view the dose distribution along any plane of the implant. Dose distributions can be projected onto the reference images using the patient anatomy as a frame of reference. The Point Dose tool can display the dose at any point in the plan simply by holding the cursor over the point of interest. The dose-volume histogram (DVH) can also be viewed for the entire dose matrix to enable this tool to be effectively used for 2D plans.



Evaluate two plans side-by-side (left), or view multiple planes (right).

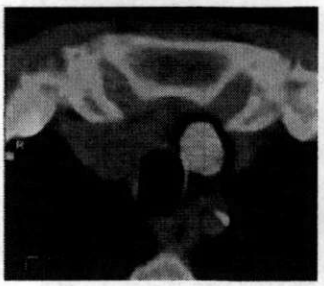
VARIAN medical systems | ONCOLOGY SYSTEMS

USA Headquarters, California
Varian Medical Systems
Palo Alto, CA
Tel: 650.424.5700
800.544.4636
Fax: 650.493.5570
www.varian.com

Headquarters Europe, Eastern Europe, Africa, Middle & Near East
Varian Medical Systems International AG
Zug, Switzerland
Tel: 41.41.749.8844
Fax: 41.41.745.5340
info.europe@varian.com

RAD7154B

7/06 (5K)



Eclipse™ Treatment Planning System

Proton Planning

Eclipse Proton Planning combines the latest in fast, accurate proton calculation algorithms with the power of the Eclipse treatment planning system to create a single system for proton, photon, electron, and brachytherapy planning. Proton plans can be combined and compared with other modalities, or used as the base dose for a boost. Eclipse supports a variety of proton beam lines as well as ocular planning. Clinicians use automated proton planning tools to easily take advantage of the unique physics of protons to reduce healthy tissue damage.

State-of-the-Art Algorithms

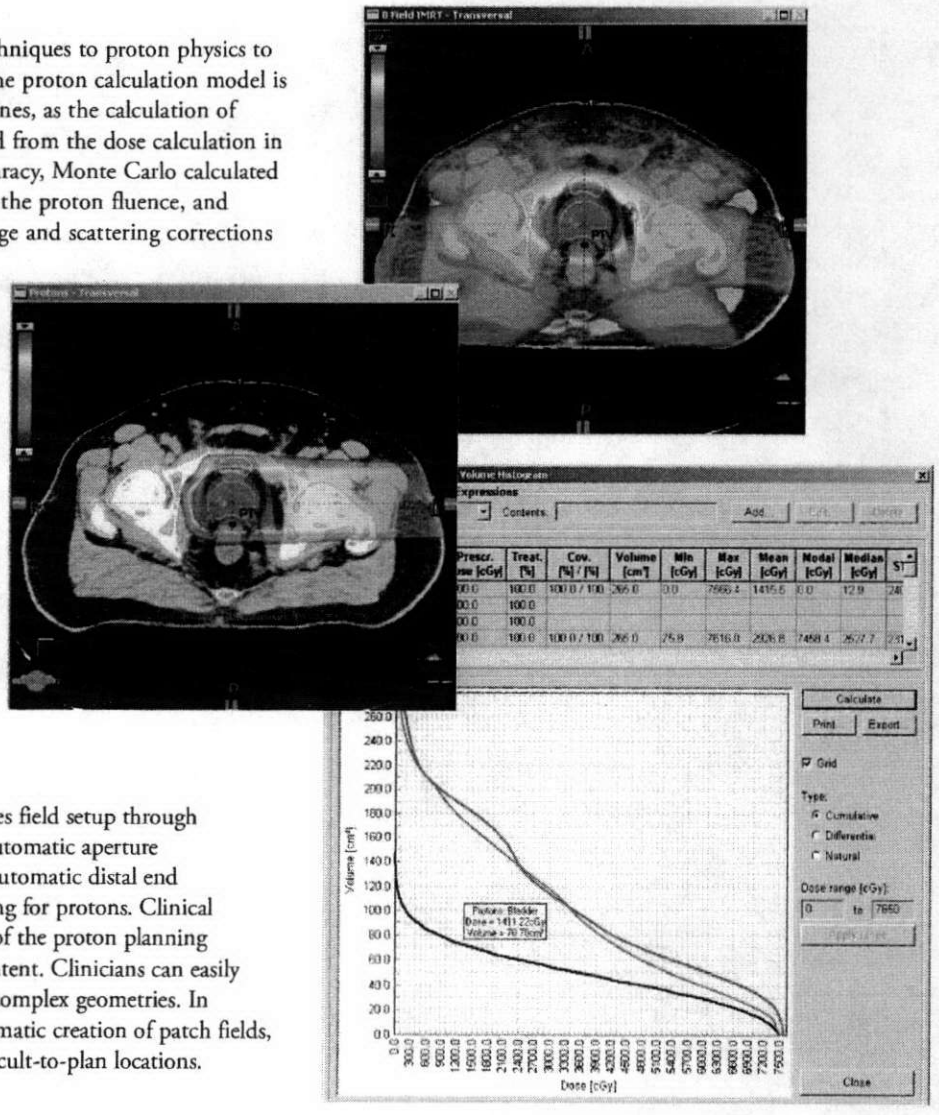
Eclipse applies Monte Carlo techniques to proton physics to derive accurate pencil beams. The proton calculation model is adaptable to a variety of beam lines, as the calculation of proton fluence in air is separated from the dose calculation in the patient. For the highest accuracy, Monte Carlo calculated dose kernels are convolved with the proton fluence, and individual dose kernels have range and scattering corrections applied to them.

Powerful Contouring Tools

Effective proton planning requires clear delineation of the tumor and critical organs. To speed the contouring process, structure sets are fully defined as part of clinical protocol templates. Eclipse has a broad range of automatic, semi-automatic, and intuitive manual tools for outlining and segmenting tumors. Powerful post-processing automatically refines the drawn structures.

Automatic Field Setup

Eclipse Proton Planning automates field setup through automatic isocenter placement, automatic aperture conformation to structures, and automatic distal end compensators to speed up planning for protons. Clinical protocol templates preload most of the proton planning parameters from the physician's intent. Clinicians can easily copy and oppose fields for more complex geometries. In addition, with the easy semi-automatic creation of patch fields, they can deliver protons into difficult-to-plan locations.



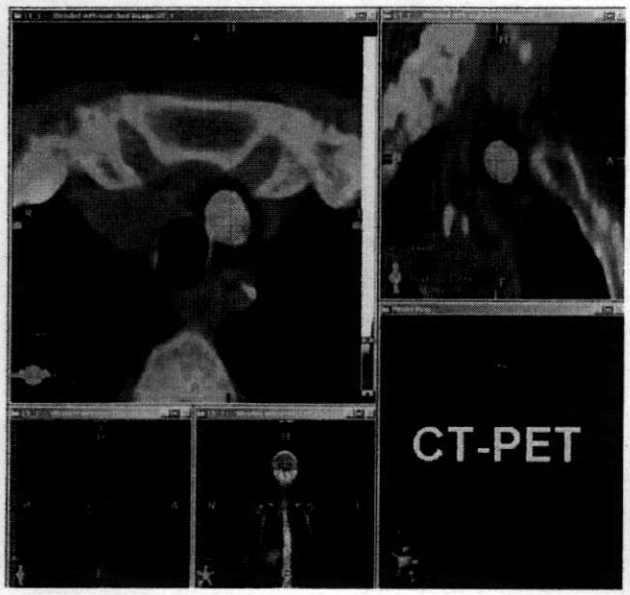
A dose-volume histogram (DVH) of the rectum shows a comparison between a single-field proton plan and an intensity-modulated radiation therapy (IMRT) plan.

Eclipse™ Treatment Planning System

Multi-Modality Imaging

Eclipse Proton Planning provides excellent visualization of tumor and critical structures. Clinicians can co-register and blend separate data sets obtained with CT, MR, or PET—a particularly important capability when considering the high dose gradients of protons.

Co-registered images also simplify contouring on any of the modalities. For example, contouring on a PET image automatically appears on the CT image.



Using a co-registered CT/PET image, multi-modality imaging permits accurate delineation of a tumor's extent.

Open System

Eclipse supports industry-standard import and export via DICOM RT, including Treatment Plan, Plan Dose, and Planar Dose. For protons, Eclipse supports DICOM RT Ion export of compensator and blocking files to milling machines for fabrication. A DICOM RT to ASCII converter in Eclipse allows the physicist to export plan data to common spreadsheets for independent evaluation.

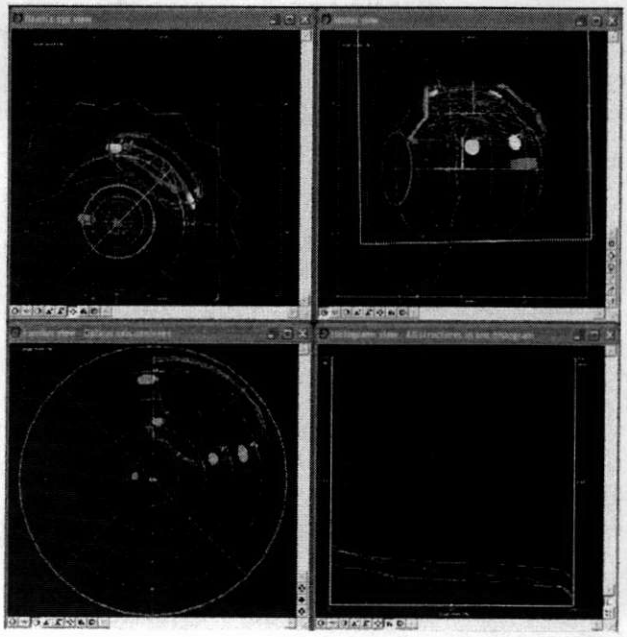
Integrated Environment

Proton plans can be combined, compared, and evaluated, along with photon, electron, and brachytherapy modalities in a single integrated Eclipse environment. This capability ensures the optimal treatment strategy for each patient.

Ocular Proton Planning

For proton treatment of ocular neoplastic tumors, Eclipse offers efficient planning software based upon geometric models for eyes shaped either spherically or elliptically. The clinician can add upper and lower eye lids to the model for either a constant rim thickness or constant lid thickness. To speed the process, Eclipse automatically conforms blocks to the tumor shape and automatically determines range definition and modulation for the optimal treatment. Other features include multiple tumor support and wedges for beam shaping. To determine the best eye position for treatment, interactive fixation light adjustment is available with real-time dose calculation and multiple histogram views. The user can create comprehensive reports that include images and simulation comparisons.

Eclipse Proton Planning for ocular lesions uses a broad beam algorithm. Lateral, distal, and proximal penumbras are taken from beam line measurements. All the penumbras are constant throughout the target and are independent of range and modulation. There is an option to add more profiles in which the distal penumbra is characterized by the range, the proximal profile by the spread-out Bragg peak (SOBP) and the lateral penumbra by the depth. Eclipse interpolates between these profiles to obtain the characteristics for the current configuration.



Eclipse Ocular Proton Planning screen capture showing a plan with a wedge used for beam shaping.

VARIAN medical systems | ONCOLOGY SYSTEMS

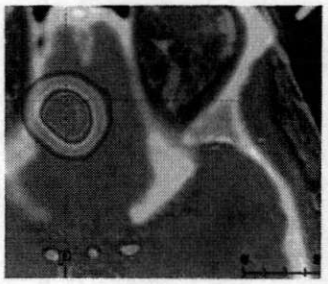
USA Headquarters, California
Varian Medical Systems
Palo Alto, CA
Tel: 650.424.5700
800.544.4636
Fax: 650.424.5520
www.varian.com

Headquarters Europe, Eastern Europe, Africa, Middle & Near East
Varian Medical Systems International AG
Zug, Switzerland
Tel: 41.41.749.8844
Fax: 41.41.759.8411
info.europe@varian.com

RAD7160B

7/06 (5K)

 2.26



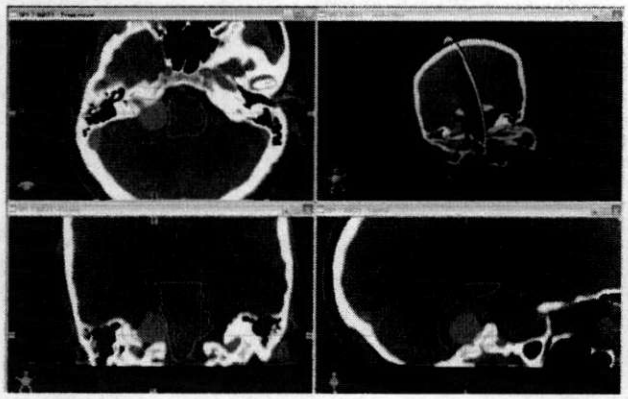
Eclipse™ Treatment Planning

Varian Stereotactic Planning

Stereotactic planning software from Varian offers a complete and comprehensive treatment planning solution for all accelerator-based stereotactic radiosurgery and radiotherapy (SRS/SRT) programs. With this software, clinicians can rapidly create highly conformal treatment plans that spare healthy tissue for delivery with cones or multileaf collimators (MLCs) from a variety of manufacturers. A combination of clinical protocols, real-time dose updates, and optimization speed the planning process for both intracranial and extracranial SRS.

Complete Stereotactic Solution

The stereotactic planning software supports all accelerator-based delivery techniques that employ stereotactic cones or multileaf collimators. Clinicians can plan single-fraction and fractionated conformal arc as well as intensity-modulated radiosurgery (IMRS) treatments for MLCs on Varian, Elekta, Siemens, BrainLab, and Mitsubishi accelerators. The software supports accessories and head frame systems from multiple manufacturers.



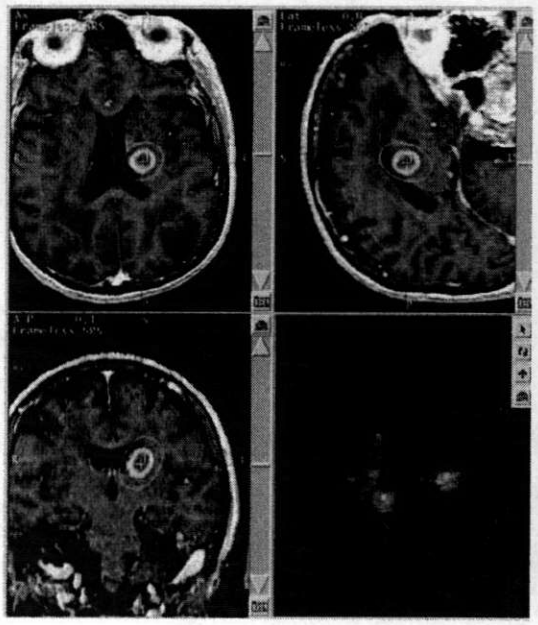
An IMRS plan for an acoustic neuroma of the brain illustrates the high degree of conformality to the lesion.

For both intracranial and extracranial SRS, Eclipse supports the entire planning process from image management to quality assurance. Automatic and manual tools rapidly import and register CT, MR, and PET images in stereotactic coordinate space. Multiple contrasting color schemes highlight the tumor and critical structures in blended images.

High-Quality Radiosurgery Plans

The clinician can specify the conformality index and an unlimited number of isocenters per plan for optimum target coverage. Real-time update of dose display in axial, sagittal, and coronal views helps the planner to quickly modify the treatment to spare critical structures. Rendering in 3D displays isodose surfaces along with the target volumes and critical structures.

The clinician can compare multiple plans and quickly select the optimum plan for a given patient. The system can synchronize side-by-side display of candidate plans so that the same view from each plan is used in evaluation. Calculation and display of dose-volume histograms (DVHs) for multiple structures from alternative plans highlight the relative benefits of each plan.



A frameless SRS treatment of a brain metastasis using cones shows excellent dose conformality and rapid dose fall-off.

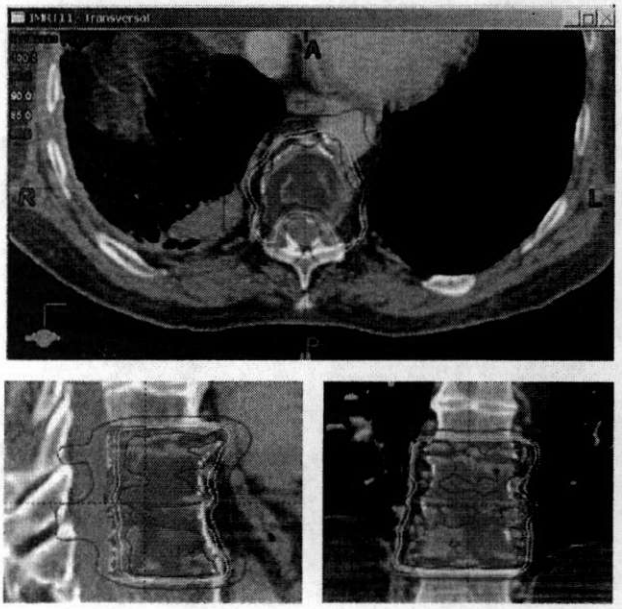
 227

Eclipse™ Treatment Planning

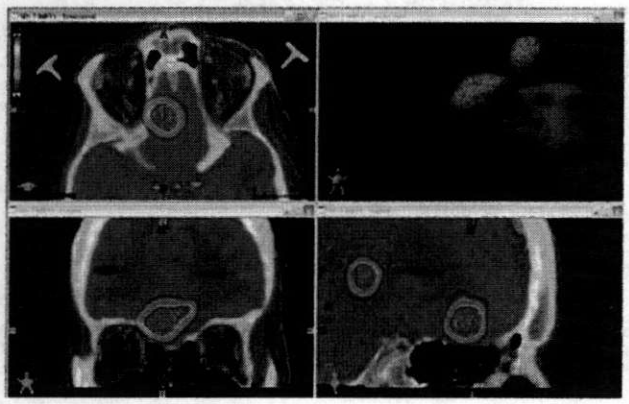
Fast Planning

Varian stereotactic planning produces highly conformal dose distributions that spare healthy tissue without requiring the clinician to draw contours. An automatic sphere-packing algorithm creates single or multiple-isocenter plans in seconds. Whenever the planner changes parameters, the system immediately updates dose distributions to allow for rapid assessment and adjustment. A normal tissue constraint eliminates the need to define structures for sparing in many IMRS cases. Use of a single isocenter to treat several lesions reduces IMRS delivery time from hours to minutes.

User-defined clinical protocol templates automate every step of the radiosurgery planning process based upon diagnosis and clinician's treatment preference. The clinical protocol templates contain structure sets, beam geometry, accessories, dose prescription, treatment plan objectives, IMRS objectives, and DVH calculation settings. Any plan may be saved as a clinical protocol, organized into user-defined folders, and applied consistently to subsequent patients. Eclipse sets up the planning parameters so the clinician can focus on developing the best plans.



Excellent sparing of the spinal cord is demonstrated with an extracranial seven-field IMRS treatment plan.



Two brain lesions are treated simultaneously with a single isocenter using IMRS, substantially reducing treatment time.

An Integrated Solution

IMRS planning gains many advantages from integration with the Eclipse™ treatment planning system. These include centralized patient data, distributed planning process, interactive IMRS optimization, logic-based and image-based automatic segmentation, thorough plan documentation and a comprehensive system of user rights. The clinician can pre-plan an IMRS treatment from any Eclipse computer in the department, including the one on the physician's desk.

Electronic approval of plans prevents unintended modification and ensures plan integrity. Eclipse reduces the time required for quality assurance of IMRS plans by calculating verification plans for either a scanned or geometric phantom, or for the PortalVision™ MV imaging system.

When part of the Varian Inspiration™ integrated oncology environment, an approved plan immediately becomes available for treatment. Eclipse uses the DICOM standard extensively to ensure that treatment plan information can be accessed by multiple vendor devices.

When Eclipse is included with the Varian Trilogy™ stereotactic system, delivering stereotactic treatments is optimized with advanced beam delivery, on-board imaging, and precise patient positioning. Isocenter coordinates can transfer to Varian's optical guidance system for accurate patient setup and position monitoring during the entire delivery. Digitally reconstructed radiographs (DRRs) generated in Eclipse for selected beam geometries can be compared to kV images from the On-Board Imager® kV imaging system (OBI) for patient anatomy-based position verification.

VARIAN medical systems | ONCOLOGY SYSTEMS

USA Headquarters, California
Varian Medical Systems
Palo Alto, CA
Tel: 650.424.5700
800.544.4636
Fax: 650.493.1349
www.varian.com

Headquarters Europe, Eastern Europe, Africa, Middle & Near East
Varian Medical Systems International AG
Zug, Switzerland
Tel: 41.41.749.8844
Fax: 41.41.759.5911
info.europe@varian.com

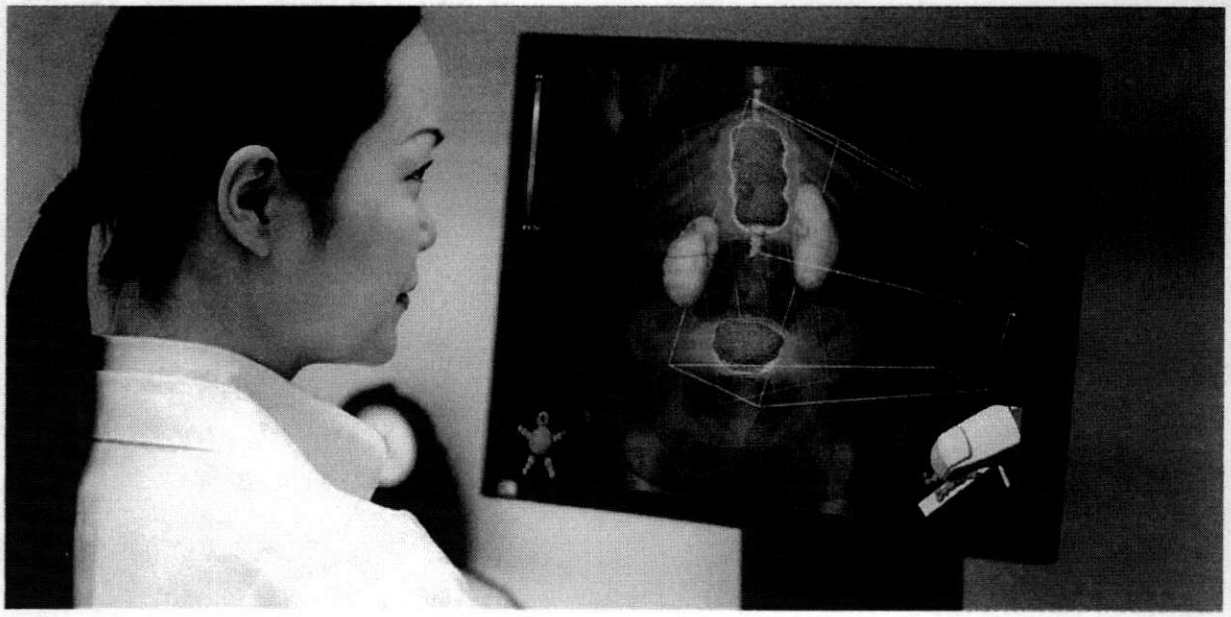
RAD7162B

7/06 (5K)

VARIAN
medical systems

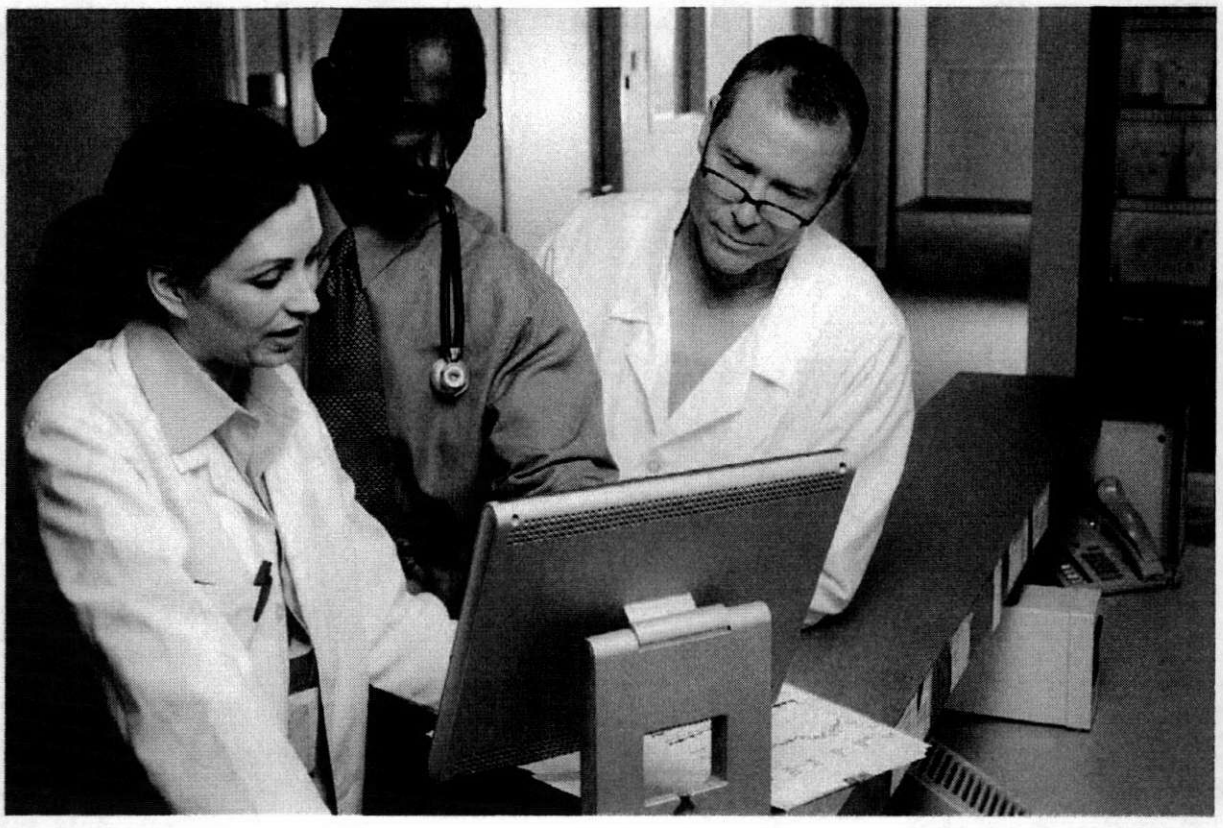
Eclipse Treatment Planning System

New Features and Enhancements for 2010



FEATURE	FUNCTION	BENEFIT	
Registration	Multimodality image registration and segmentation	Increased planning versatility, PET contouring, and enhanced target volume delineation	Included
Portal Dosimetry for RapidArc® Radiotherapy Technology	Pre-treatment verification of RapidArc plans	Ability to split RapidArc treatment fields into verification sub-arcs	Included*
Leaf Motion Calculator	Tracking jaws based on MLC aperture	Reduced leakage, support of step-and-shoot and multiple static segment IMRT delivery	Included
Proton Therapy	New spot list editor; improved setup field handling; room's eye view imaging; improved beam line editor	Custom designed modulated scanning	Included
RapidArc Radiotherapy Technology	Intuitive arc beam geometry tool, progressive resolution algorithm, and simple collision detection	Faster treatment field setup, reduced optimization time, and efficient motion calculation	Purchasable option
Smart Segmentation® Automatic Contouring Utility	Fully automatic segmentation	Improved contouring efficiency and consistency	Purchasable option
SmartAdapt	Deformable registration; toolset for tracking and adapting to interfractional changes during treatment	Quickly visualize and address differences between the planning CT and CBCT images	Purchasable option
Eclipse Application Programming Interface (API)	Ability for clinicians to implement their own dose calculation and optimization algorithms within Eclipse	Increased planning flexibility	Purchasable option

**Included if clinical site already has RapidArc and Portal Dosimetry licenses*



Registration

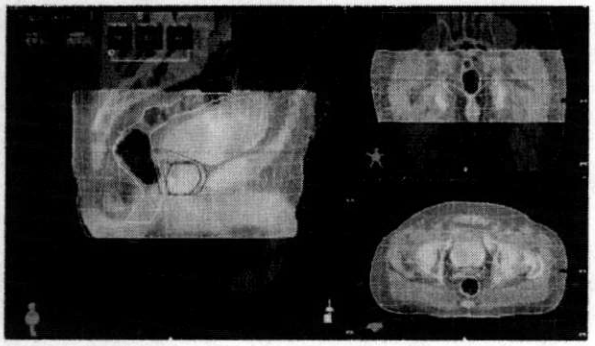
The Eclipse™ treatment planning system features an improved multimodality image registration and segmentation workspace for PET, CT, MR, as well as CBCT scans. A selection of several automatic rigid registration algorithms, including point-based matching, provides you with the flexibility to plan both simple and complex cases and allows integration of many aspects of morphology and function into the contouring process. Manual rigid registration, registration review, and approval are also included. PET and relative intensity-based contouring capabilities enable a variety of uses for biological imaging in radiation oncology.

Smart Segmentation

The Smart Segmentation automatic contouring utility features enhanced segmentation of bony structures of the pelvis, the external head contour, and the thoracic spinal canal. The newly added segmentation of the seminal vesicles better addresses the need for prostate cancer patient segmentation and has the potential to positively influence the segmentation of other organs such as the rectum. To support optimal image-guided workflow and better patient positioning, gold markers segmentation has been introduced. Improved artifact correction, support of basic segmentation for different body positions, the option to customize the dimension of the segmented structures and other new features allow you to use this application for a wider range of patients. In addition, Smart Segmentation enables several segmentation services to be supported in a multi-site environment.

SmartAdapt

The SmartAdapt toolset enables you to easily track and adapt to interfractional changes in anatomy in your patient throughout the treatment. Using the interactive deformable registration and segmentation tools, you can quickly visualize and address differences between the planning CT and CBCT images. You can automatically deform and propagate initial contours to match the current anatomy, and edit or fine tune the changes using a variety of 2D and 3D contour editing features. Sophisticated techniques such as blending, color mapping, difference rendering, and deformation-specific information allow a thorough review and approval of your registration results. A statistics tool allows tracking of relevant changes and provides you with crucial information for deciding about a potential reoptimization. Adapting treatments to meet your clinical goals and to improve patient outcomes can be made easier with SmartAdapt.



Display of deformation grid and deformation distance information after deformable registration of two CT scans of a patient with prostate cancer.

Image courtesy of Dr. J. Wu, Duke University, Durham, NC

RapidArc enhancements

The next generation of RapidArc radiotherapy technology provides tools which simplify the set up of the RapidArc treatment field. The new intuitive arc Geometry Tool will suggest the number of arcs and isocenters based on target position and size, and improves efficiency during the planning process. New features, including the use of automatic normal tissue objective, mean dose objective, and intermediate dose calculation, will increase the quality of your treatment plans and make it possible to treat a wider range of tumor sites.

Progressive-Resolution Optimizer Algorithm

The enhanced Progressive-Resolution Optimizer Algorithm (PRO) available in 2010 significantly reduces the optimization time for RapidArc treatment plans. To optimally achieve radiation goals and parameters given by radiation oncologists, the improved PRO uses calculated dose distribution done with AAA as a reference dose when restarting the optimization.

Leaf Motion Calculator

The leaf motion calculator algorithm is available only for the TrueBeam™ system and is fully integrated in the Eclipse distributed calculation framework (DCF) to support both sliding window and multiple static segment IMRT delivery.

Portal Dosimetry for RapidArc

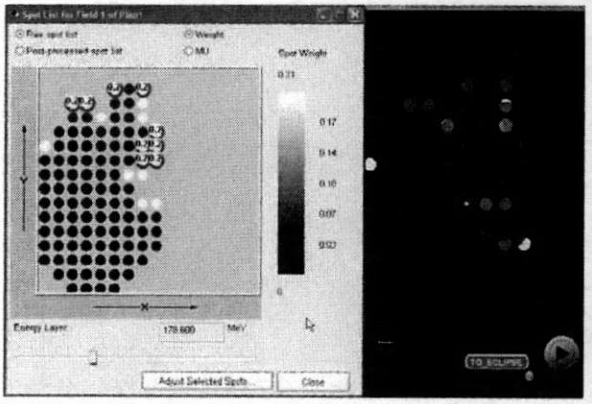
Get more value out of your PortalVision™ MV imaging system and eliminate the need to purchase additional QA equipment just for RapidArc. The Eclipse treatment planning system provides tools which allow you to split your RapidArc treatment field into verification sub-arcs for detailed investigation of small arc segments. You can use Portal Dosimetry for RapidArc pre-treatment QA – giving you the same fast and efficient process, and the same high-resolution QA measurements for RapidArc.

Eclipse Treatment Planning System

Proton therapy

The Eclipse treatment planning system offers the next generation of proton therapy. Eclipse continues to support advanced proton delivery capabilities. Now you can tailor your proton treatments using the spot list editor. This new feature allows you to edit the spot list for custom designed modulated scanning.

The proton planning features are fully integrated with advanced hardware and software components, providing a range of configuration options specifically tailored to each site's clinical objectives.



The spot list editor displays values and spots to be edited.

Key components include:

- Enhanced beam-line editing capability, including user specification of spread out Bragg peak (SOBP) and range
- Room-based imaging, including gantry pitch angles
- Enhanced setup field handling
- Beam spot editor for pencil-beam scanning

VARIAN
medical systems
A partner for life

USA Headquarters, California
Varian Medical Systems
Palo Alto, CA
Tel: 650.424.5700
800.544.4636
Fax: 650.493.5833
www.varian.com

Headquarters Europe, Eastern Europe, Africa, Middle & Near East
Varian Medical Systems International AG
Zug, Switzerland
Tel: 41.41.749.8844
Fax: 41.41.740.3340
info.europe@varian.com

RAD 10118

7/10 (200)

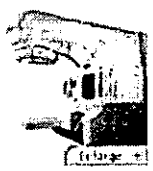
VARIAN medical systems | A partner for life

Oncology | Radiation Oncology
Trilogy | Overview

- **ARIA**
  Information System
- **Eclipse**
  Treatment Planning
- **Acuity**
  Verification & Simulation
- **Trilogy**
  Treatment Delivery
- **Clinac**
  Treatment Delivery
- **Upgrades**

Overview

Trilogy – Power. Precision. Versatility.



The new Trilogy system adds more powerful advanced motion management capabilities.

Complete with Gated RapidArc and a new open motion management interface, these tools expand the options for treating moving tumors.

Gated RapidArc utilizes the RPM system to monitor and adjust for tumor motion during a RapidArc treatment.

The open interface can also be used with third party companies to connect their devices to Varian accelerators for the purpose of monitoring motion. This interface further demonstrates the open architecture design of the Trilogy system and provides support to third party companies to monitor tumor motion in real time.

The power of Trilogy yields treatment times that are shorter, making the experience more comfortable for the patient.

The precision of Trilogy allows you to spare healthy tissues to an extent that was unimaginable only a few years ago.

The versatility of Trilogy enables treatment of a wide variety of patients using a single machine.

All In One. Best In One.

Varian offers the complete package in Trilogy:

- Choose from a broad range of external beam therapies, including 3D CRT, IMRT, IGRT or DART using the Trilogy system.
- Multiple dose rate options – up to 1000mu/min for efficient SRS delivery.
- 2D and 3D KV image guidance for higher quality imaging at lower doses.
- Full 360° range of treatment delivery angles with positional couch angles.
- Stereotactic frame or frameless immobilization for patient positioning - treat any area of the body.
- Real-time Position Management™ (RPM) system - for gating perfectly timed beam delivery with minimal margins.
- Gated RapidArc with advanced motion management and open third-party interface.
- Portal Dosimetry IMRT treatment delivery verification.
- Dynamic high resolution MLC for exquisite beam sculpting.
- Delivery verification and quality assurance in Argus Linac and Argus IMRT quality assurance software.

Clinical benefits:

- Highest dose rate for shorter sessions.
- Faster treatment times.
- Tight isocenter alignments on all three axes. Targets the smallest lesions.
- Rapid on-board imaging. Reposition patients quickly and accurately.
- Cone-Beam CT, fine tune patient set ups with ultra-precise CT scans.

Varian is not trying to make you follow our process, we provide you with the number one products that will support your processes.

The Trilogy system is the first in a new generation of cancer care systems. A versatile system optimized for multiple forms of treatment, from radiation therapy to radiosurgery.

The versatile Trilogy system delivers 3D conformal radiotherapy, IMRT, stereotactic radiosurgery, fractionated stereotactic radiation therapy, and intensity-modulated radiosurgery for cancer and neurosurgical treatment.

Contact Varian Oncology

Varian Oncology Headquarters
Tel: 1.650.424.5700
Contact Us

Events

- **08/08/2010 - 08/13/2010**
  CAARI 21st International Conference on the Application of Accelerators in Research & Industry
- **08/08/2010 - 08/21/2010**
  15th Brazilian Congress of Medical Physics
- **08/11/2010 - 08/13/2010**
  MOGA - Medical Oncological Group of Australia
- **08/18/2010 - 08/21/2010**
  World Cancer Congress (International Union Against Cancer)

VARIAN medical systems | A partner for life

Oncology | Radiation Oncology
Trilogy | Treatment Techniques

- **ARIA**
  Information System
- **Eclipse**
  Treatment Planning
- Acuity
  Verification & Simulation
- **Trilogy**
  Treatment Delivery
- Clinac
  Treatment Delivery
- Upgrades

Treatment Techniques

The most treatment choices for the best personalized care.

Only Varian Medical Systems offers you so many imaging, planning and treatment options. Our flexible technologies–the finest and most automated in the industry–make it possible for you to customize treatment to suit any of the clinical circumstances you may encounter.

RapidArc™ Radiotherapy Technology – one revolution is all it takes

The innovative RapidArc radiotherapy technology on a Trilogy® system represents the latest evolution of cancer treatment technology, setting new benchmarks for speed, precision and patient comfort.

RapidArc uses a unique algorithm that provides unprecedented treatment delivery control. As a result, treatment plans that excel in covering target goals while sparing critical structures can be developed and performed faster than ever before.

RapidArc is efficient. Its single gantry rotation speeds treatment delivery so clinicians can develop treatments that take one-half to one-eighth the time of conventional IMRT treatments—just two minutes in many cases. A RapidArc treatment may also result in less radiation leakage and scatter, so peripheral tissues receive a lower overall dose.

Varian IGRT — The future is in motion. We'll be there to help you manage it.

Deliver IMRT without holding your breath.

While IMRT is a way of conforming dose to the target, IGRT makes it possible to account for respiration and be assured that the target is in the same position every day.

With Varian IGRT, innovative radiographic, fluoroscopic and cone beam CT modes are integrated with automated repositioning and motion management visualization software, so you can verify treatments that are synchronized with respiration.

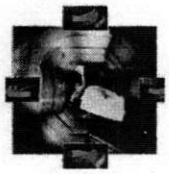
IGRT lets you obtain high resolution, three dimensional images to pinpoint tumor sites, adjust patient positioning when necessary, and complete a treatment–all within the standard treatment time slot. With IGRT, even sophisticated IMRT and stereotactic IMRT become more efficient.

Clinical features of IGRT from Varian

- Choose from the widest array of integrated in-room imaging tools to meet any clinical need–cone beam CT, radiographic, fluoroscopic, MV imaging, optical imaging and ultrasound
- Adapt therapies at treatment with advanced IGRT motion management and remote couch motion
- Manage organ motion for patient free breathing, breath-hold or gating with the Real-time Position Management™ (RPM) system

- Share data across the health enterprise with our comprehensive oncology information system
- Get better image guidance. Powerful treatment planning enables clinicians to combine CBCT. PET, CT and MR images

Varian IMRT – Better technology at every step.



IMRT is fast becoming the standard of care. Not all tumors are alike. Varian IMRT optimize all forms of IMRT delivery, from low resolution to high resolution. Choose the resolution you need according to each clinical situation.

IMRT "paints" dose to the tumor with pinpoint precision, while sparing healthy normal tissue. Its performance has been so impressive that clinicians around the world are using it in the treatment of nearly every type of solid tumor. Increasingly, clinics are treating patients with Varian IMRT and nearly every new linear accelerator ordered from Varian is equipped to deliver IMRT. It has proven to be an extremely effective way to treat cancer and improve patients' lives during treatment and beyond.

Clinical benefits of high resolution IMRT

- Minimizes hot spots
- Improves target inhomogeneity
- Provides detailed dose painting to the target
- Sculpts dose around critical structures more effectively
- Allows treatments to occur in conventional 10 to 15 minute time slots
- Dose resolution–up to 500 segments per field
- Spatial resolution: 2.5mm-5mm

Contact Varian Oncology

Varian Oncology Headquarters
Tel: 1.650.424.5700
Contact Us

Events

- **08/08/2010 - 08/13/2010**
  CAARI 21st International Conference on the Application of Accelerators in Research & Industry
- **08/08/2010 - 08/21/2010**
  15th Brazilian Congress of Medical Physics
- **08/11/2010 - 08/13/2010**
  MOGA - Medical Oncological Group of Australia
- **08/18/2010 - 08/21/2010**
  World Cancer Congress (International Union Against Cancer)

2856

VARIAN medical systems | A partner for life

Oncology | Radiation Oncology
Trilogy | Treatment Delivery Technology

- **ARIA**
  Information System
- **Eclipse**
  Treatment Planning
- Acuity
  Verification & Simulation
- **Trilogy**
  Treatment Delivery
- Clinac
  Treatment Delivery
- Upgrades

Treatment Delivery Technology

A new direction in cancer care

Trilogy combines new image-guidance technology that lets you know exactly where the tumor is at the moment of treatment, with the most focused and powerful radiation beam available.

The result is faster and highly precise treatments.

On-board imager (OBI) automated image registration tools calculate the couch offsets to automated patient repositioning.

Highly conformal dose distributions

Trilogy includes intensity-modulated beam delivery, which results in dose distributions that match the 3D shape of the target volume, including targets with concave and complex shapes.

Versatile

Trilogy has been optimized for radiosurgery and radiotherapy and is designed to raise the standard of care while lowering the cost of treatment.

Both neurosurgeons and radiation oncologists can use Trilogy to treat their patients, without having to buy two separate, expensive machines.

Power

- Trilogy has a more tightly focused beam and can help deliver doses significantly faster than conventional accelerators
- The result is
  - reduced effects of tumor motion during treatment
  - shortened treatment times
  - enhanced patient comfort

The benefit of stereotactic approaches are generally realized with smaller lesions. As better diagnostic tools result in earlier detection, when the tumor is still quite small, stereotactic treatments are likely to play an increasing role in radiation oncology.

The Trilogy system incorporates Varian's Dynamic Image Guided Radiation Therapy (IGRT) accessories for improved treatment precision, whether using conventional, conformal, or stereotactic approaches to radiation therapy.

2057

Contact Varian Oncology

Varian Oncology Headquarters
Tel: 1.650.424.5700
Contact Us

Events

- **08/08/2010 - 08/13/2010**
  CAARI 21st International Conference on the Application of Accelerators in Research & Industry
- **08/08/2010 - 08/21/2010**
  15th Brazilian Congress of Medical Physics
- **08/11/2010 - 08/13/2010**
  MOGA - Medical Oncological Group of Australia
- **08/18/2010 - 08/21/2010**
  World Cancer Congress (International Union Against Cancer)

VARIAN medical systems | A partner for life

Oncology | Radiation Oncology
Trilogy Accelerator

- **ARIA**
  Information System
- **Eclipse**
  Treatment Planning
- **Acuity**
  Verification & Simulation
- **Trilogy**
  Treatment Delivery
- **Clinac**
  Treatment Delivery
- Upgrades

Trilogy Accelerator

Meeting a higher set of standards

Trilogy is the first practical, clinically-viable system for delivering all forms of external-beam radiation therapy. It enables doctors to choose and use the most appropriate treatment modality for treating cancer in the body or the head and neck, and to deliver the full spectrum of treatments, all on one machine in a single room.

Built on the solid foundation of the Clinac iX platform the Trilogy is a fast, versatile and comprehensive delivery system. Trilogy represents a cost-effective means of expanding the patient case mix and treatment modality repertoire within a radiation oncology department.

Delivery system components

- The Trilogy configuration of the Clinac iX accelerator
- Exact IGRT couch
- Millennium 120-leaf MLC with dynamic MLC capability
- RPM respiratory gating system
- LaserGuard collision detection system
- 4D Integrated Treatment Console

The Trilogy accelerator has all the capabilities of a Clinac 21EX or 23EX plus the following new features:

- Stereotactic mode (6MV beam, up to 1000MU/min dose rate, up to 6000 MU/field total dose, up to a maximum field size of 15cm x 15cm, and up to 60MU/deg dose rate for arc-based treatments)
- 0.5mm radius isocenter for gantry and collimator axes
- 0.75mm radius isocenter for all three rotational axes (gantry, collimator and couch)
- Remote couch motion
  - Small translations and small rotations (less than 2 cm and 2 degrees) for fine-tuned setup corrections
  - Large rotations for remotely sequencing between non-coplanar arcs

Tight isocenter alignments on all three axes. Targets the smallest lesions.

Trilogy leads in beam accuracy for treating increasingly smaller tumors anywhere in the body. On two axes (the gantry and the collimator) Trilogy's isocentric accuracy measures 0.5mm or less. Add the third rotational axis (the couch) and the isocenter radius is 0.75mm or less, the tightest alignment available in a dual energy system.

Maximize efficiency: a common approach to integration

Trilogy takes advantage of the common approach to integration taken for granted in today's Varian products. This gives you the advantage of information presented in a consistent way, available when and where you want it throughout Varian's Aria information oncology system. In addition, training overhead is reduced as staff are familiar and confident using all aspects of the equipment. Greater accuracy is achieved as the same patient positioning systems can be used throughout the planning and treatment process. Ease of use is a priority.

Contact Varian Oncology

Varian Oncology Headquarters
Tel: 1.650.424.5700
Contact Us

Events

- **09/12/2010 - 09/16/2010**
  ESTRO 29
- **09/15/2010**
  Emerging Technologies Clinical Forum
- **09/22/2010**
  Emerging Technologies Clinical Forum
- **09/22/2010 - 09/25/2010**
  CARO - ASRO Canadian Association of Radiation Oncology 34th Annual Meeting

VARIAN medical systems | A partner for life

Oncology | Radiation Oncology
On-Board Imager (OBI) – confidence in tumor targeting

- ARIA
  Information System
- **Eclipse**
  Treatment Planning
- Acuity
  Verification & Simulation
- **Trilogy**
  Treatment Delivery
- Clinac
  Treatment Delivery
- Upgrades

On-Board Imager (OBI) – confidence in tumor targeting

The Varian On-Board Imager® kV imaging system (OBI) is standard on the Trilogy linear accelerator and makes dynamic targeting IGRT more efficient and convenient.

Treat what you have planned - providing the ability to deliver more accurate treatment with confidence and repeatability.

Whether you want to image the patient daily or weekly OBI has the capability to define flexible imaging protocols that meet your needs.

The OBI kV imaging system delivers improved tumor targeting using high resolution, low dose digital imaging in the treatment room. Users can confidently manage patient and target movement- both before and during treatments. A choice of imaging modalities - 2D radiographic, fluoroscopic or 3D Cone beam CT imaging. The use of kV imaging means lower patient dose and better image quality than megavoltage imaging.

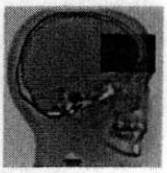
You need the best image information to help you make fast accurate decisions when comparing reference images to set up images prior to treatment – this information is what you base your decision on whether to continue to treat or make a change.

Clinical challenges

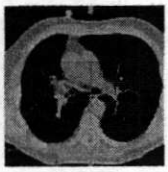
- An easy to use and efficient system is paramount for IGRT to be a real clinical solution that can be implemented for all patients.
- Versatility to enable standard protocols to be supported but also that individual patients needs can be supported to.
- Image quality needs to be good enough to be able to make clinical decisions quickly and accurately before treatment delivery.
- Acceptable dose.

Clinical features

- High quality means you can quickly see set up and anatomical differences, therefore enabling quick and accurate decisions before treatment.
- Low dose imaging means that it is acceptable to do daily imaging of required.
- Choice of imaging modalities that can be used whenever you need to using preset protocols or manual selection.
- Easy-to-use user interface with automated comparison tools quickly show differences in patient positioning.
- Automated extension and retraction of OBI hardware does not require operator intervention in the room.



The planning and delivery of tightly defined dose distributions in three dimensions is a challenge that has been solved. Yet aligning the treatment beam to the target volume remains a key challenge to improving outcomes because patients, tumors and normal tissues move. The OBI provides the tools to manage both interfraction motion (changes in position caused by day-to day set up conditions) and intrafraction motion (changes in position during a treatment session because of normal respiratory and organ motion).

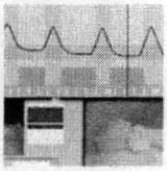


To minimize both types of motion, the OBI introduces pretreatment online imaging to the clinical process.

Managing interfraction motion with ease. Fast. Reliable. Automation.

The OBI identifies the patients position at the time of treatment so that small couch adjustments can position the patient optimally. The system has two modes of operation – radiographic repositioning using fast reliable automated tools for anatomical registration or radio opaque marker registration, and cone beam CT repositioning. The OBI is designed for convenience so that all types in the patient repositioning process can be performed remotely – saving time and simplifying the process.

Managing Intrafraction motion



Tumors that move because of respiratory motion may be more effectively treated using the Real-time Position Management™ (RPM) system. Gating automatically turns the treatment beam on and off at intervals synchronised with the patient's respiratory pattern, effectively freezing the moving target.

Robotic Technology – geared to performance, accuracy and image quality

The OBI is mounted on the treatment machine via robotically controlled arms which operate along three axes of motion so that they can be positioned for the best possible view of the tumor. Positioning is all automated.

By using the robotic technology and control software to position the OBI and patient couch, the complete system offers the automation, speed, and flexibility needed to make IGRT treatment clinically practical for thousands of cancer patients. All functionality is supported in a fully integrated environment.

Contact Varian Oncology

Varian Oncology Headquarters
Tel: 1.650.424.5700
Contact Us

Events

- **09/12/2010 - 09/16/2010**
  ESTRO 29
- **09/15/2010**
  Emerging Technologies Clinical Forum
- **09/22/2010**
  Emerging Technologies Clinical Forum
- **09/22/2010 - 09/25/2010**
  CARO - ASRO Canadian Association of Radiation Oncology 24th Annual Meeting

- Annual Reports
- Stock Information
- Financials
- Events and Webcasts
- SEC Filings
- XBRL Filings
- Corporate Governance
- Shareholder Services
- Investor Contacts

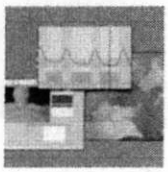
VARIAN medical systems | A partner for life

Oncology | Radiation Oncology
Trilogy | RPM respiratory gating

- ARIA
  Information System
- **Eclipse**
  Treatment Planning
- Acuity
  Verification & Simulation
- **Trilogy**
  Treatment Delivery
- Clinac
  Treatment Delivery
- Upgrades

Real-time Position Management™ (RPM) System

Respiration synchronised imaging and treatment



The Real-time Position Management™ (RPM) system is a non-invasive, video based system that allows for clean imaging and treatment of lung, breast, and upper abdominal sites. The RPM system is accurate, easy to use and fast. It is comfortable for the patient and accommodates both breath hold and free breathing protocols.

The RPM system allows you to correlate where the tumor is in relation to the patients respiratory cycle. Using an infrared tracking camera and a reflective marker, the system measures the patients respiratory pattern and range of motion and displays them as a waveform. The gating thresholds are set when the tumor is in the desired portion of the respiratory cycle. These thresholds determine when the gating system turns the treatment beam on and off.

The RPM system facilitates the treatment of the lungs, liver and pancreas, and helps minimise dose to the heart in breast treatments. The system is designed to be used anywhere you encounter the effects of respiratory motion. It also provides clean images for planning, so that you can more clearly visualise the target with fewer of the image artefacts associated with respiratory motion.

Key features:

- Only the Varian RPM system has clinically proven interfaces to CT and CT/PET scanners.
- Lightweight marker ensures patient comfort.
- Predictive filter – patented by Varian monitors and predicts the patient's breathing pattern and can account for patient's coughing or changes from the predicted breathing pattern.
- Eclipse has immediate access to gated verification images from the On-Board Imager system during the course of the treatment. Recalculation of dose distributions may be necessary.
- Superior Clinac® design means <100millisecond beam on times with accurate gated dose rates and uncompromised beam flatness and symmetry.
- Pre-treatment verification can be performed by the on-board imager fluoroscopy mode.
- Gated radiographic image acquisition using the On-Board Imager during the correct phase of the respiratory cycle.
- Gated port film acquisition – acquires MV images during the correct phase of the respiratory cycle.
- Automated process of image acquisition and gated treatments.
- After treatment a record of the gating trace and image window can be stored for later analysis.
- Treatment beam gating integrated with dynamic MLC leaf motion.

Phase Gating and Amplitude Gating

The RPM system supports both methods:

Phase based gating allows automatic gating of image acquisition and treatment delivery based on the same phase of the patients respiratory cycle. This image shows gating on the exhalation phase.

Amplitude based gating allows automatic gating based on the absolute position of the marker block on the patients thorax or abdomen, regardless of the phases in the patients respiratory cycle. This image shows gating at maximum inhalation.

Patient evaluation and post plan verification

Gating can be used on Acuity for patient evaluation and suitability for gating.

The Acuity system allows efficient pre-treatment verification of gating thresholds and anticipated gated treatment delivery before setting up the patient on

the accelerator for treatment. If adjustments are required for the gating parameters, these can be done in advance of the treatment delivery. Acuity supports both 2D and 3D positioning verification and fluoroscopic evaluation of gating thresholds for post plan verification.

Clinical versatility:

- All clinical breathing protocols, including free breathing are supported.
- Allows you to monitor respiration without sacrificing accuracy and patient comfort.
- The system can be used for retrospective and prospective gating.
- Gating allows for patient specific treatment margins, rather than population based margins.
- Gating may facilitate in dose escalation.
- Accurate tumor tracking allows maximum dose to tumor and minimum dose to normal tissue
- The RPM system is in routine use worldwide.

From simulation to treatment the RPM system allows you to accurately monitor and compensate for tumor movement.

Contact Varian Oncology

Varian Oncology Headquarters
Tel: 1.650.424.5700
Contact Us

Events

- **09/12/2010 - 09/16/2010**
  ESTRO 29
- **09/15/2010**
  Emerging Technologies Clinical Forum
- **09/22/2010**
  Emerging Technologies Clinical Forum
- **09/22/2010 - 09/25/2010**
  CARO - ASRO Canadian Association of Radiation Oncology 24th Annual Meeting

- Oncology Training
- Oncology Support
- X-Ray Products Support
- Security and Inspection

VARIAN medical systems | A partner for life

Oncology | Radiation Oncology
Benefits of Trilogy

- **ARIA**
  Information System
- **Eclipse**
  Treatment Planning
- **Acuity**
  Verification & Simulation
- **Trilogy**
  Treatment Delivery
- **Clinac**
  Treatment Delivery
- **Upgrades**

Benefits of Trilogy

The Trilogy is the world's first image-guided radiation therapy system optimized for both conventional and stereotactic approaches to treating cancer. The versatile Trilogy™ system delivers IGRT as standard, and can be used to deliver 3D conformal radiotherapy, IMRT, stereotactic radiosurgery, fractionated stereotactic radiation therapy, and intensity-modulated radiosurgery for cancer and neurosurgical treatment.

Administrator

Trilogy provides a flexible and efficient treatment delivery system with state of the art treatment capabilities. This advanced functionality in one elegant machine helps improve patient care and increase the number of patients treated. This translates to a high profile reputation for the facility, increased revenues, and an accelerated return on investment.

The Trilogy linear accelerator enables busy clinics to offer more advanced treatment to more patients. Trilogy can deliver IGRT as standard, in addition to 3D conformal radiotherapy, IMRT, stereotactic radiosurgery, fractionated stereotactic radiation therapy, and intensity-modulated radiosurgery for cancer and neurosurgical treatment.

In addition, the unprecedented flexibility of the Trilogy comes with the best customer support service in the industry and market leading reliability. With decades of experience as a world leader, Varian is committed to ensuring your machines are always up and running and delivering the highest quality treatment care.

Physician/oncologist

The versatility of the Trilogy enables any area of the body to be treated using the latest and most effective treatments available. This includes advanced radiosurgery treatments. With full under couch access and multi-treatment modalities, the flexibility of the Trilogy helps clinicians offer treatments from a range of techniques suited to the needs of each patient.

In the case of IMRT, this includes:

- Step and shoot
- Sliding window
- Small to large fields
- Coplanar or non-coplanar fields
- Radical or palliative plans

The Trilogy is designed to deliver a high dose rate for fast treatment times and to help ensure effective hypofractionation. As a result, more patients can be treated quickly and accurately.

Physicist

The machine and dose stability of the Trilogy makes IMRT, IGRT and IMIGRT treatment delivery easy and effective. The Trilogy is designed so the beam remains consistent and can be quickly turned on and off. This high degree of accuracy and reliability enables gating and other advanced treatment techniques and ensures that the dose output is the same according to treatment.

The Trilogy also features streamlined matching of machine dosimetry. This means that machines can be beam matched across the department, so patients can be easily and quickly transferred from machine to machine if required.

Therapist

The Trilogy is designed with the operator in mind so it is easy to use and includes a large number of automated processes and safety mechanisms as standard features. This level of automation provides a logical working process for improved efficiency and safety. As a result, therapists can concentrate on explaining the process to the patient, rather than operating the machine.

In a busy facility where therapists frequently rotate around the department, commonality between products—for instance familiar controls, user interfaces and hand pendants—means less training and the ability to quickly transition through the treatment process. Trilogy supports a streamlined clinical environment with an integrated approach designed to help save time and confusion. The result? A safe treatment environment where more patients can be treated reliably and effectively.

Read more about the benefits of Trilogy at leading treatment centers:

- Lung cancer: Edward Cancer Center
- Intestinal cancer: Emory University School of Medicine
- Lung and other cancers: University of Maryland
- Lung cancers: UPMC

Contact Varian Oncology

Varian Oncology Headquarters
Tel: 1.650.424.5700
Contact Us

Events

- **08/08/2010 - 08/13/2010**
  CAARI 21st International Conference on the Application of Accelerators in Research & Industry
- **08/08/2010 - 08/21/2010**
  15th Brazilian Congress of Medical Physics
- **08/11/2010 - 08/13/2010**
  MOGA - Medical Oncological Group of Australia
- **08/18/2010 - 08/21/2010**
  World Cancer Congress (International Union Against Cancer)

VARIAN medical systems | A partner for life

Oncology | Radiation Oncology
Trilogy | Product Resources

- **ARIA**
  Information System
- **Eclipse**
  Treatment Planning
- **Acuity**
  Verification & Simulation
- **Trilogy**
  Treatment Delivery
- **Clinac**
  Treatment Delivery
- **Upgrades**

Trilogy | Product Resources

In support of the Varian Trilogy community, this section provides links to user's groups, our product support helpdesk, and a comprehensive list of related materials, including brochures, product catalogs, and demos.

Trilogy | MyVarian

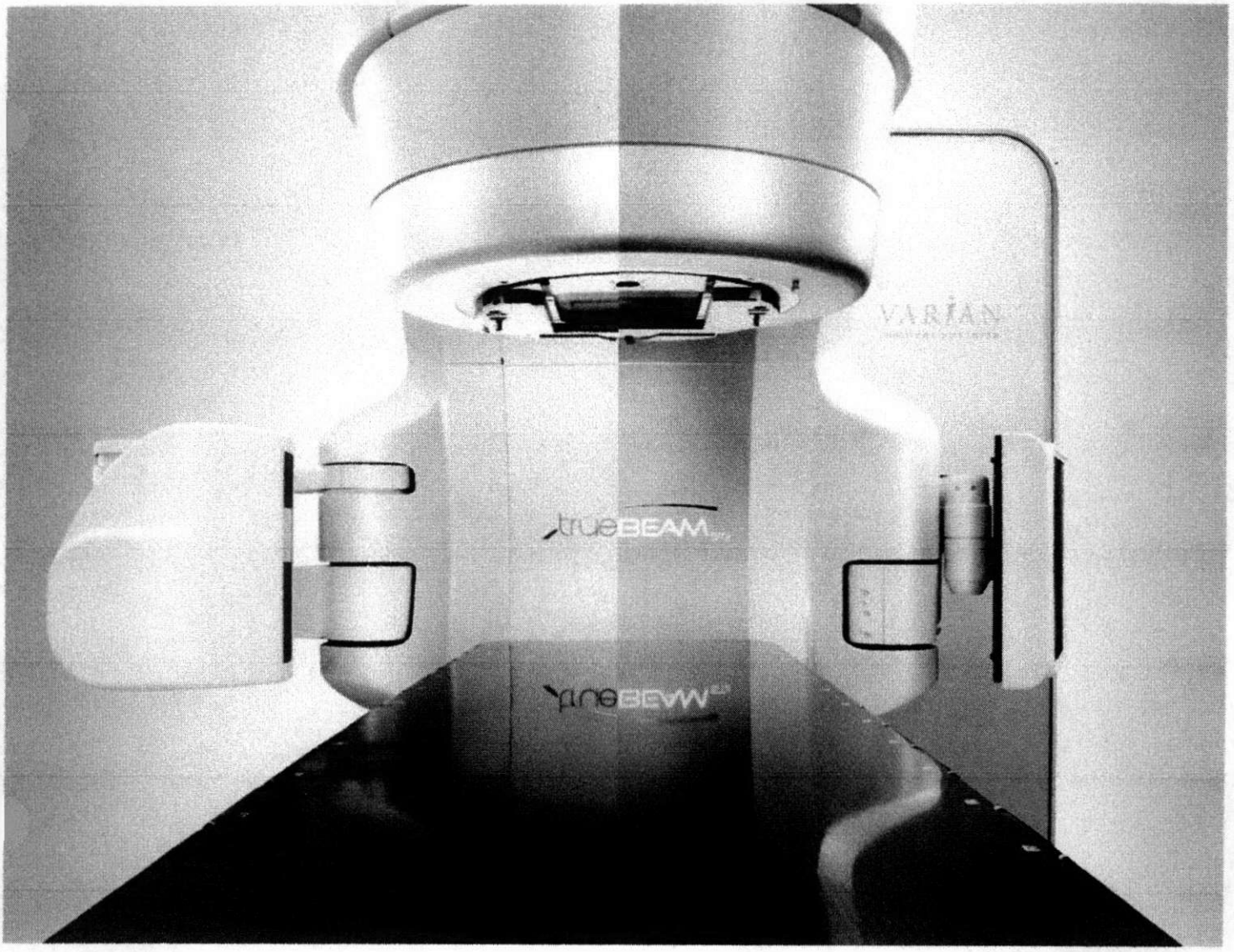
Access your MyVarian account for online Help Desk assistance, product documentation updates, and up-to-date news about Varian products.

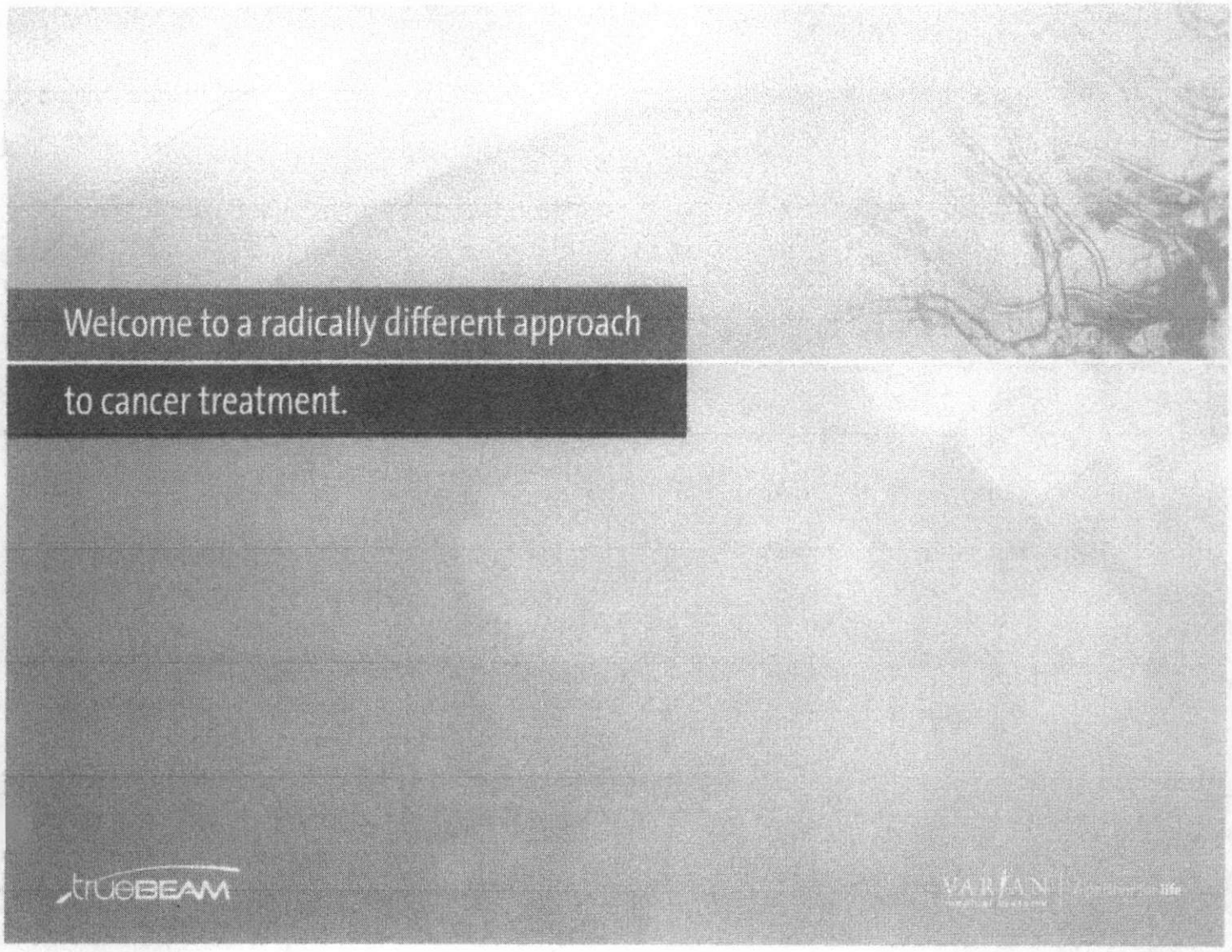
Contact Varian Oncology

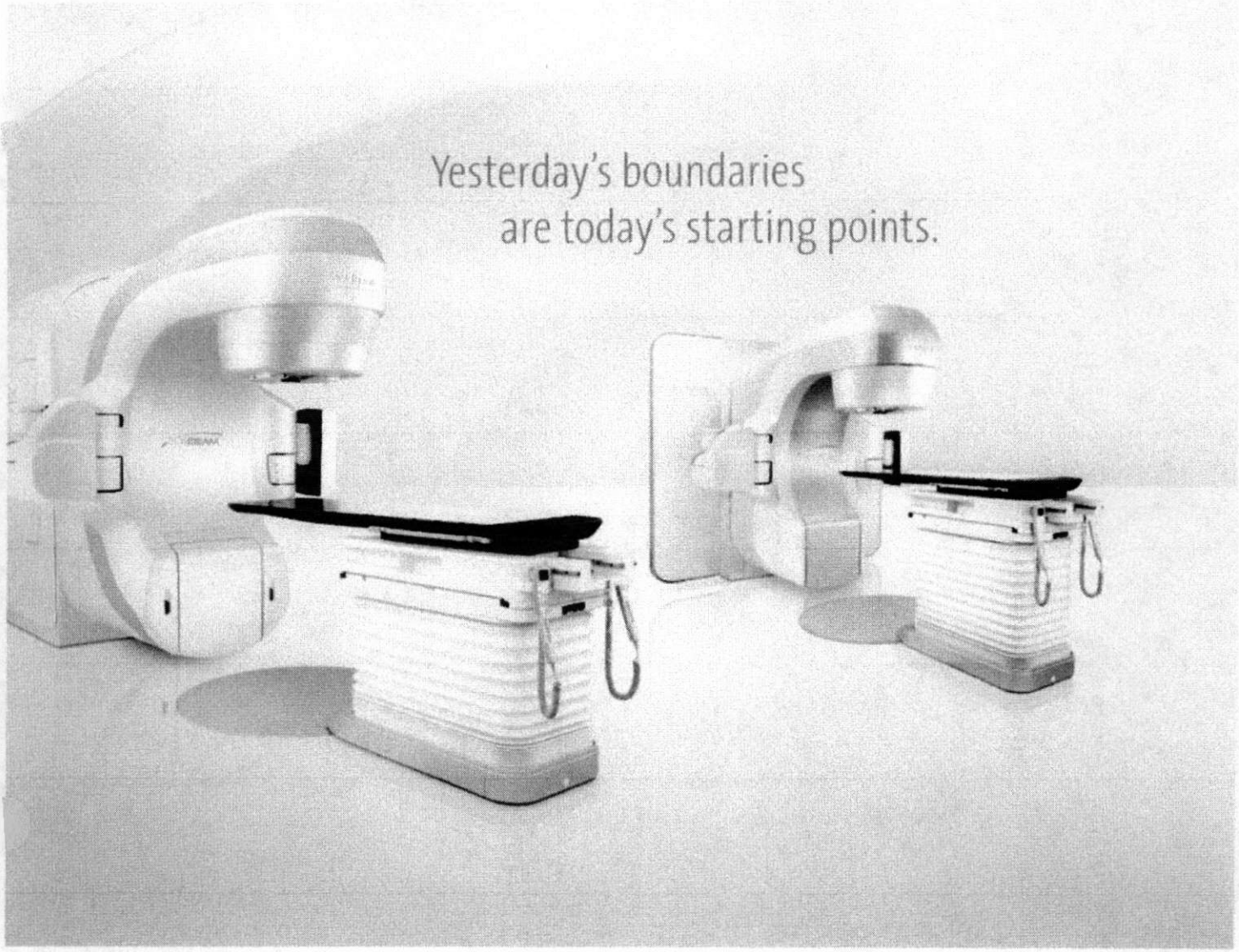
Varian Oncology Headquarters
Tel: 1.650.424.5700
Contact Us

Events

- **08/08/2010 - 08/13/2010**
  CAARI 21st International Conference on the Application of Accelerators in Research & Industry
- **08/08/2010 - 08/21/2010**
  15th Brazilian Congress of Medical Physics
- **08/11/2010 - 08/13/2010**
  MOGA - Medical Oncological Group of Australia
- **08/18/2010 - 08/21/2010**
  World Cancer Congress (International Union Against Cancer)







TRUEBEAM

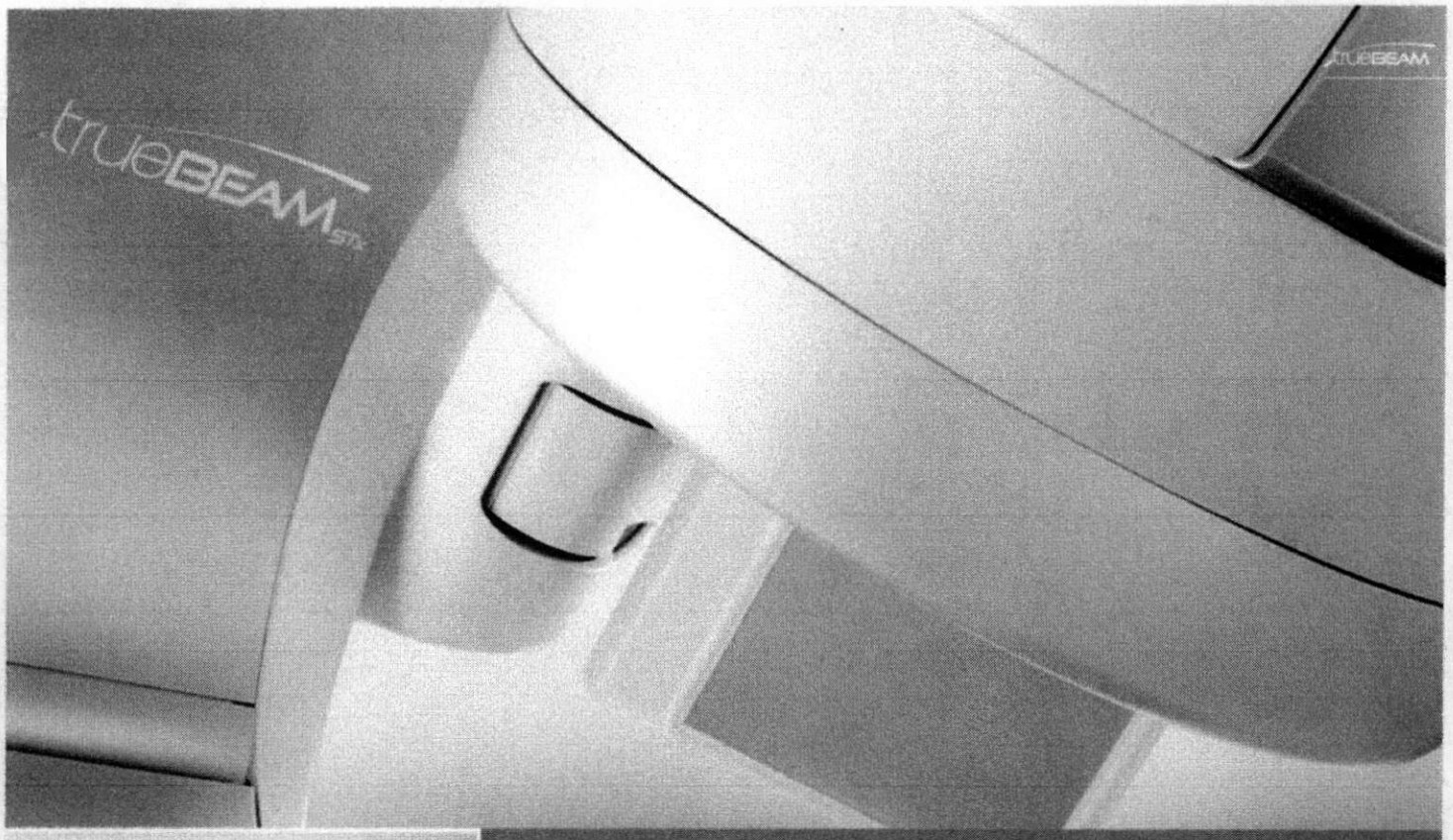
Introducing

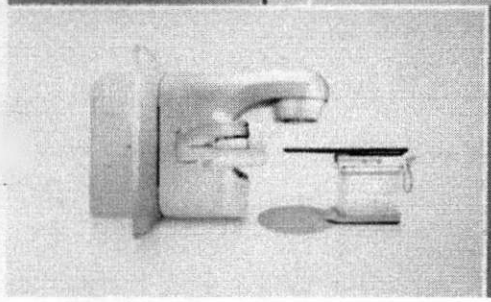
the TrueBeam™ system for radiotherapy and the TrueBeam™ STx system for radiosurgery

Break the boundaries of today's treatment technologies with a system designed to target and treat cancer with pinpoint accuracy. A system whose precision is measured in increments of less than a millimeter. A radically new solution that is faster, is more powerful and advances the opportunity for new treatment options. Designed on a sophisticated, synchronized architecture, you can now navigate the complexities of cancer care with confidence. Developed from the ground up to optimize both radiotherapy and radiosurgery, TrueBeam and TrueBeam STx are innovative, intelligent and intuitive.

Welcome to a radically different future for you and your patients.

The revolutionary TrueBeam system was engineered with a new sophisticated architecture so that advancing innovation and unlocking new treatments can become a reality. It is a system where imaging and delivery are exquisitely integrated, providing superior control for progressive treatment techniques. Now you can customize your radiotherapy treatment program to your clinical needs and make a difference with the game-changing capabilities of the TrueBeam system.

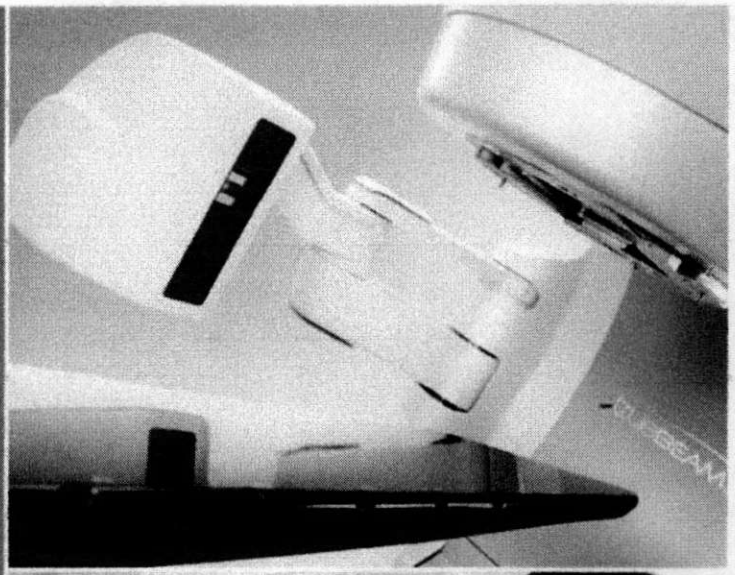


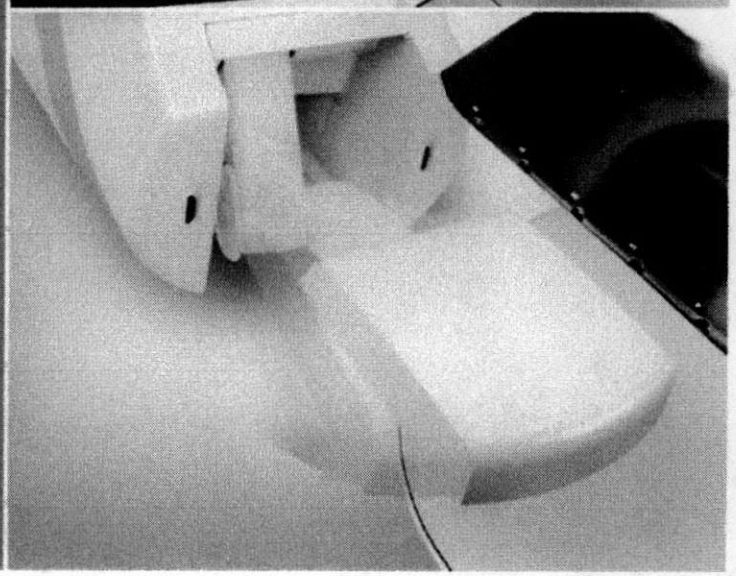


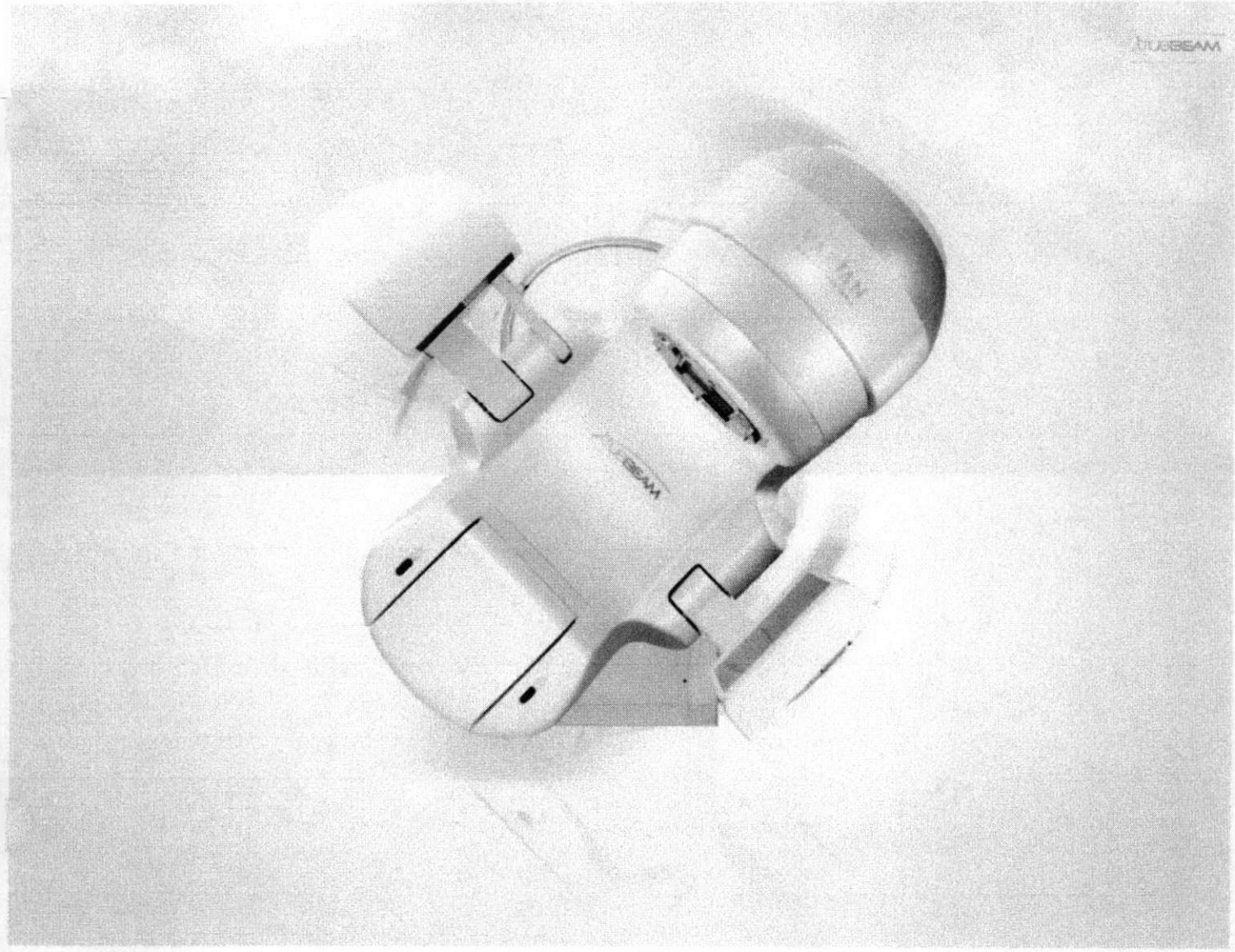
TrueBeam enables a new way to collaborate using Developer Mode. In a protected, non-clinical environment, oncology thought leaders interested in pioneering new treatment and imaging techniques can aspire to go beyond today's clinical norms.

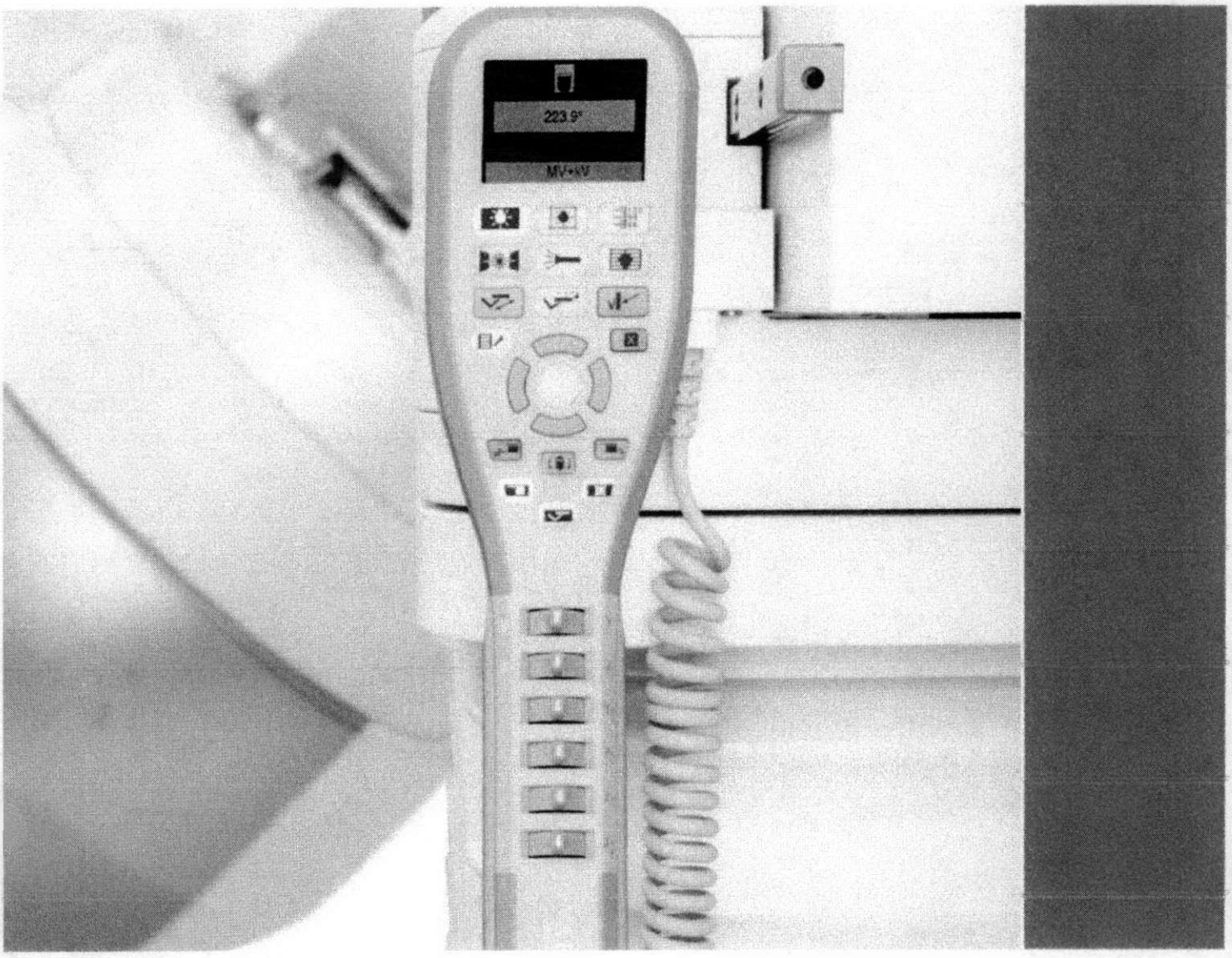
Intelligence that inspires excellence.

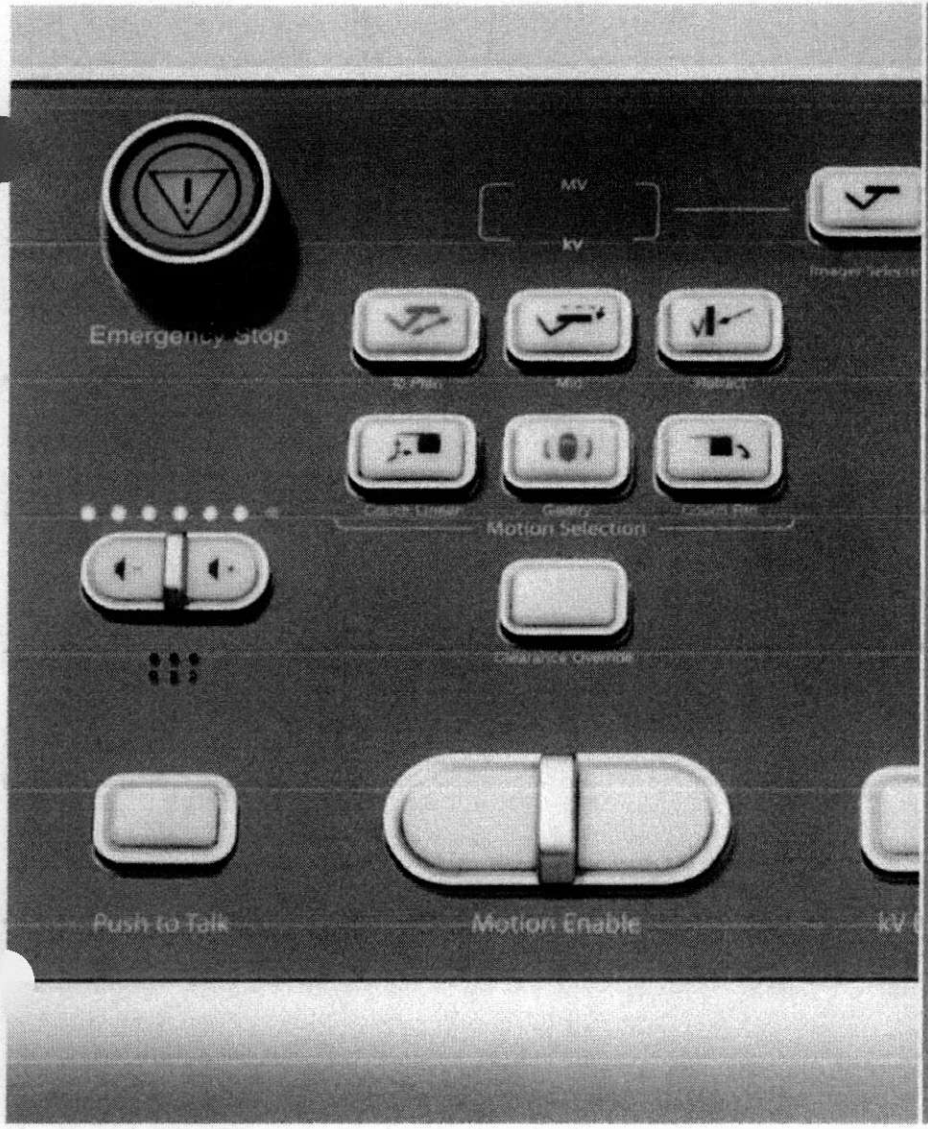
The inspiring TrueBeam system is highly intelligent. It choreographs, in real time, an asynchronous network for treatments that require fast, flexible timing capabilities. It seamlessly integrates respiratory gating, imaging and treatment techniques. The precision of the TrueBeam system is measured in increments of less than a millimeter, with mechanical and dose accuracy coming together at a true isocenter. With dose rates that are 40–140% higher than current standards, you can treat more patients faster. With an abundance of robust tools, this radical system encourages new levels of clinical excellence.







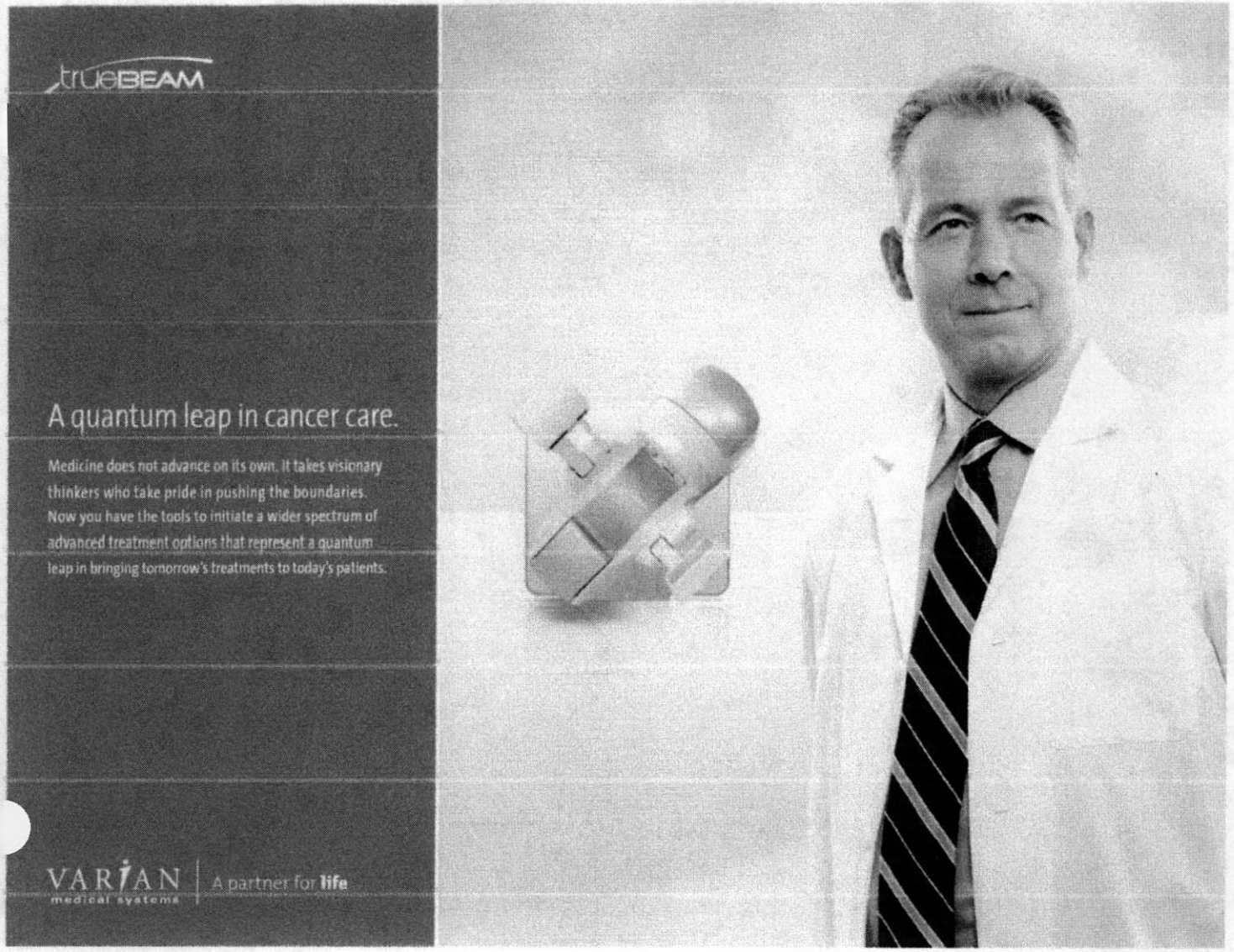


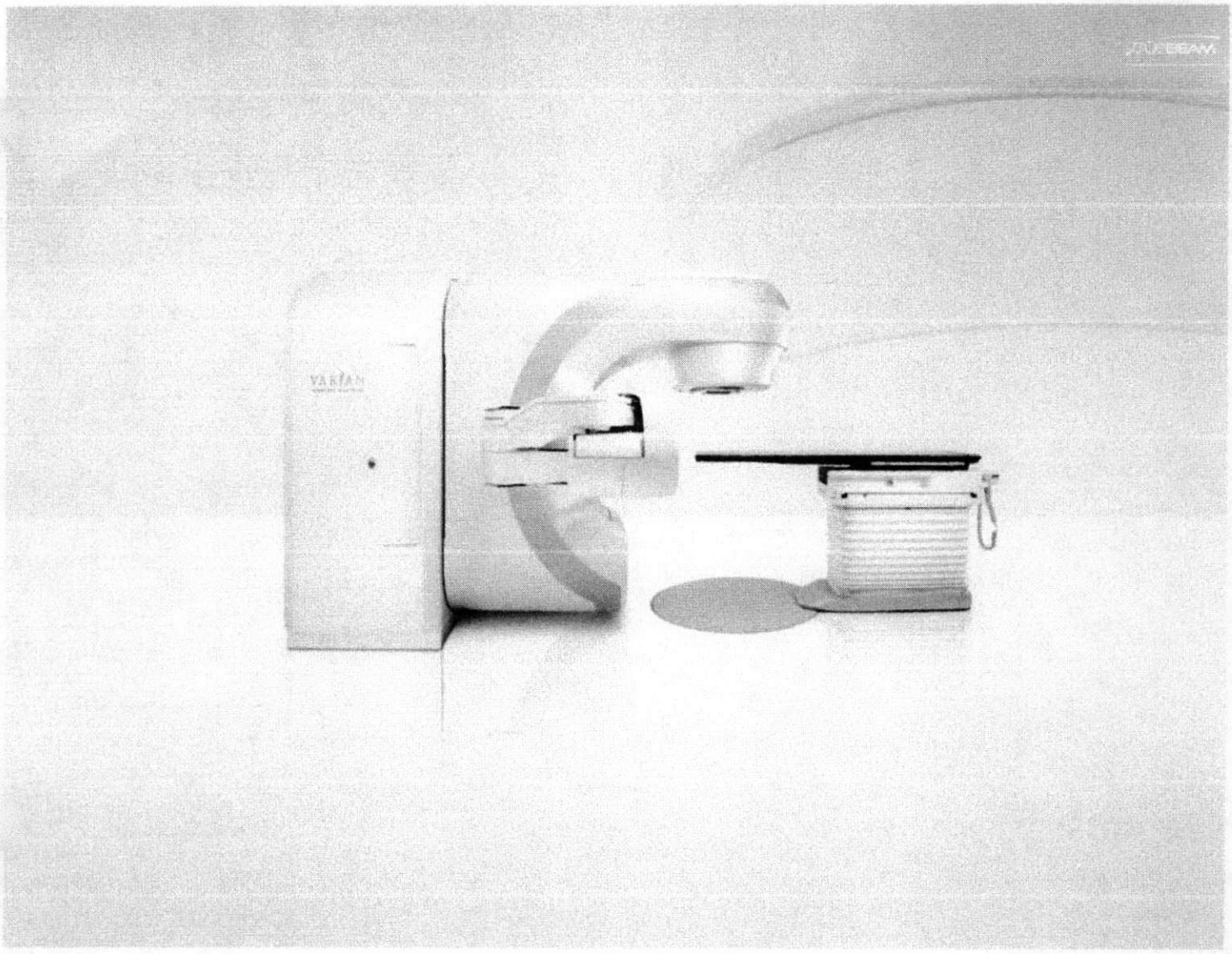


TrueBeam

It's simply intuitive.

Exceedingly advanced as it is, TrueBeam is fundamentally unique in its intuitive simplicity. Imaging and treatment processes have been automated and streamlined for improved performance. A superior workflow with features such as one-button image acquisition and full automation of beam delivery make treatments 50% faster. A simple and straightforward system operation allows therapists to make the patient the focal point of their attention.





VARIAN | A partner for life
medical systems

USA Headquarters

USA Regional Offices

New Jersey

Belgium

Germany

Italy

Spain/Portugal

China

Brazil

Georgia

European Headquarters

Finland

India

Netherlands

UK/Ireland

Japan

Australian Headquarters

Illinois

Austria

France

Varian Medical Systems India Pvt Ltd.

Scandinavia

Asian Headquarters

Latin American Headquarters

VIEWRAY INCORPORATED	Title: **Algorithm Engine Unit Test Verification Report**	(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

[redacted]

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

(b)(4), Test Report, Verification Report, Software, Test Protocols

SECTION 8

Reference Documents for the VIEWRAY™ TREATMENT PLANNING and DELIVERY SYSTEM SOFTWARE

References for ViewRay™ 510(k) Submission Sections and Attachment 2E-1 (SRS RQ-0006)

[1] Derek L G Hill, Philipp G Batchelor, Mark Holden and David J Hawkes, Medical Image Registration, Phys. Med. Biol. 46 (2001) R1–R45)
[1a] Zeng, Q. (2008). Diffusion Weighted Magnetic Resonance Image Analysis and Medical Image Registration, Ph.D. thesis, Gainesville, FL: University of Florida
[2] J.L. Schreiner, C.P. Joshi, J. Darko, A. Kerr, G. Salomons, S. Dhanesar. The role of Cobalt-60 in modern radiation therapy: Dose delivery and image guidance. Journal of Medical Physics, Vol. 34, No. 3, (2009) 133-136
[3] E J Adams and A P Warrington, A comparison between Cobalt-60 and linear accelerator-based treatment plans for conformal and intensity-modulated radiotherapy. The British Journal of Radiology, 81 (2008), 304–310
[4] I.J. Chetty, B. Curran, J.E. Cygler, J.J. DeMarco, G. Ezzell, B.A. Faddegon, I. Kawrakow, P.J. Keall, H. Liu, C.-M. C. Ma, D.W.O. Rogers, J. Seuntjens, D. Sheikh-Bagheri, and J.V. Siebers, Report of the AAPM Task Group No. 105: Issues associated with clinical implementation of Monte Carlo-based photon and electron external beam treatment planning, Med. Phys. 34(12) 2007 pp 4818-4853
[5] Khoo VS, Joon DL. New developments in MRI for target volume delineation in radiotherapy. Br J Radiol. 2006 Sep;79 Spec No 1:S2-15
[6] Brunt JN. Computed Tomography-Magnetic Resonance Image Registration in Radiotherapy Treatment Planning. Clin Oncol (R Coll Radiol). 2010 Jul 29. [Epub ahead of print]PMID: 20674300
[7] Derek L G Hill, Philipp G Batchelor, Mark Holden and David J Hawkes, Medical Image Registration, Phys. Med. Biol. 46 (2001) R1–R45)
[8] C.T. Kelley, Iterative Methods for Optimization, SIAM Philadelphia 1999 pp 39-40
[9] Lu W, Chen ML, Olivera GH, Ruchala KJ, Mackie TR. Fast free-form deformable registration via calculus of variations. Phys Med Biol. 2004 Jul 21; 49(14):3067-87
[10] Zeng Q., Y. Chen: Accurate Inverse Consistent Non-Rigid Image Registration and Its Application on Automatic Re-contouring. ISBRA 2008, LNBI 4983, pp. 293–304, 2008. Springer-Verlag Berlin Heidelberg 2008
[11] W.H. Press, S.A. Teukolsky, W.T. Vetterling, B.P. Flannery, Numerical Recipes in C the Art of Scientific Computing, 2nd Ed., Cambridge University Press, New York pp 544-550
[12] van den Boomgard, R, and R. van Balen, "Methods for Fast Morphological Image Transforms Using Bitmapped Images," Computer Vision, Graphics, and Image Processing: Graphical Models and Image Processing, Vol. 54, Number 3, pp. 254-258, May 1992
[13] W.E. Lorenson and H.E. Cline, Marching Cubes: A High Resolution 3d Surface Construction Algorithm, Computer Graphics, Volume 21, Number 4, July 1987, pp 163-169
[14] C. Fox, H.E. Romeijn, B. Lynch, Men C, D.M. Aleman, and J.F. Dempsey. Comparative analysis of ^{60}Co intensity-modulated radiation therapy. Phys Med Biol. 53(12) (2008)

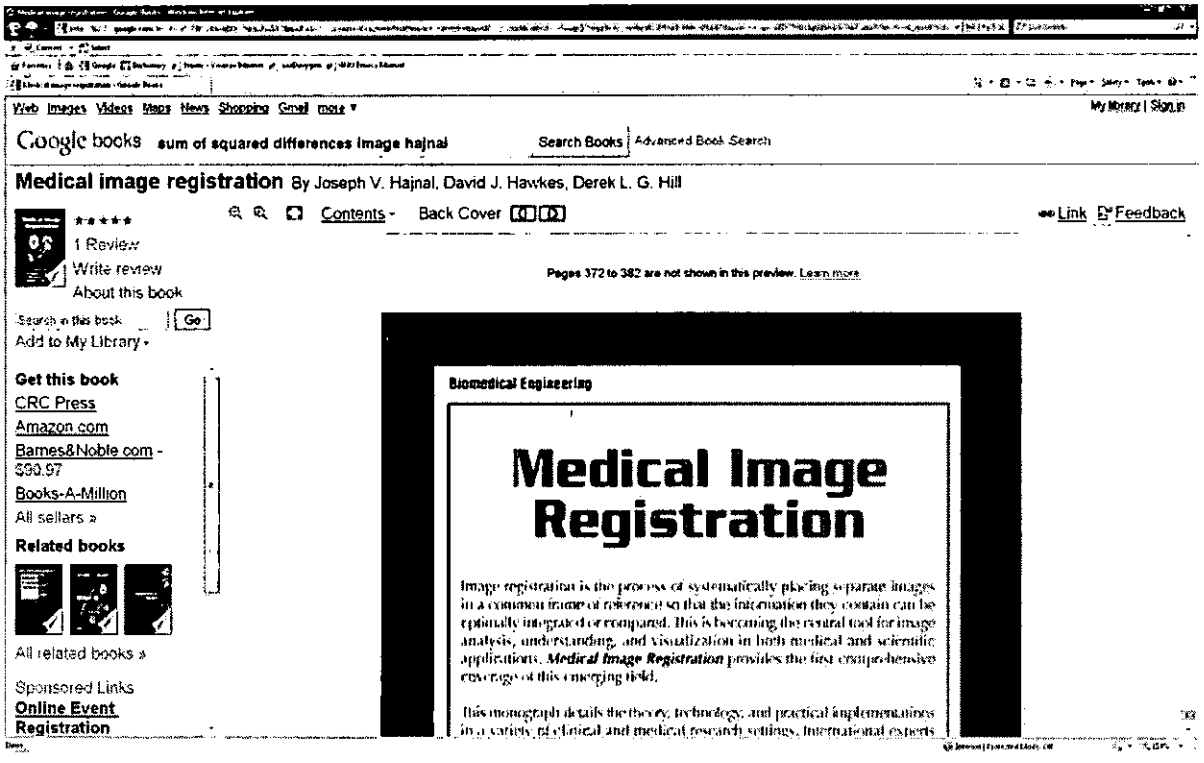
3175-88. Epub 2008 May 27
[15] Glassner, Graphic Gems, AP Professional, pp. 548-550, 1990
[16] J.L. Mott, A. Kandel, and T.P. Baker, Discrete Mathematics for Computer Scientist and Mathematicians, 2nd Ed., Prentice Hall, New Jersey, 1986
[17] A. Ahnesjo and M.M. Aspradakis, Dose calculations for external photon beams in radiotherapy, Phys. Med. Biol. 44 (1999) R99–R155)
[18] I.J. Chetty, B. Curran, J.E. Cygler, J.J. DeMarco, G. Ezzell, B.A. Faddegon, I. Kawrakow, P.J. Keall, H. Liu, C.-M. C. Ma, D.W.O. Rogers, J. Seuntjens, D. Sheikh-Bagheri, and J.V. Siebers, Report of the AAPM Task Group No. 105: Issues associated with clinical implementation of Monte Carlo-based photon and electron external beam treatment planning, Med. Phys. 34(12) 2007 pp 4818-4853
[19] J.F. Dempsey, Jonathan G. Li, H.E. Romeijn, Daniel A. Low, Jatinder R. Palta. A Fourier analysis of the dose grid resolution required for accurate fluence map optimization for IMRT treatment planning. Med. Phys. 32(2) (2005) 380-388
[20] H.E. Romeijn, R.K. Ahuja, J.F. Dempsey, A. Kumar, and J.G. Li. A novel linear programming approach to fluence map optimization for intensity modulated radiation therapy treatment planning. Phys. Med. Biol. 48(7) (2003) 3521-3542
[21] C. Fox, H.E. Romeijn, and J.F. Dempsey. Fast voxel and polygon ray-tracing algorithms in intensity modulated radiation therapy treatment planning. Medical Physics , 33 (5), pp. 1364-1371
[22] J. Sempau, S.J. Wilderman, and A. F. Bielajew. DPM, a fast, accurate Monte Carlo code optimized for photon and electron radiotherapy treatment planning dose calculations. Phys. Med. Biol., 45:2263 – 2291, 2000
[23] Chetty IJ, Charland PM, Tyagi N, McShan DL, Fraass BA, Bielajew AF., Photon beam relative dose validation of the DPM Monte Carlo code in lung-equivalent media, Med Phys. 2003 Apr;30(4):563-73
[24] Salvat F, Fernandez-Varea J M, Baro J and Sempau J 1996 PENELOPE, an algorithm and computer code for Monte Carlo simulation of electron-photon showers Ciemat (Centro de Investigaciones Energ´eticas, Medioambientales y Tecnol´ogicas) Technical Report no 799
[25] R.A. Yanez and J.F. Dempsey, A modified version of the DPM Monte Carlo code for radiotherapy treatment planning inside of uniform magnetic fields, unpublished ViewRay Report, 2007
[26] D. Jette. Magnetic fields with photon beams: Dose calculation using electron multiple scattering theory. Med. Phys., 27(8):1705, 2000
[27] Z. Li and J.F. Williamson, Volume-based geometric modeling for radiation transport calculations, Med. Phys. 19(3) 1992 pp 667-677
[28] H.E. Romeijn, J.F. Dempsey, and Jonathan G. Li. A unifying framework for multi-criteria fluence map optimization models. Phys. Med. Biol. 49(21) (2004) 1991-2013
[29] Kamath S., Sahni, S., Palta, J., Ranka, S., & Li, J. (2004). Optimal leaf sequencing with elimination of tongue-and-groove underdosage. Physics in Medicine and Biology , N7-N19

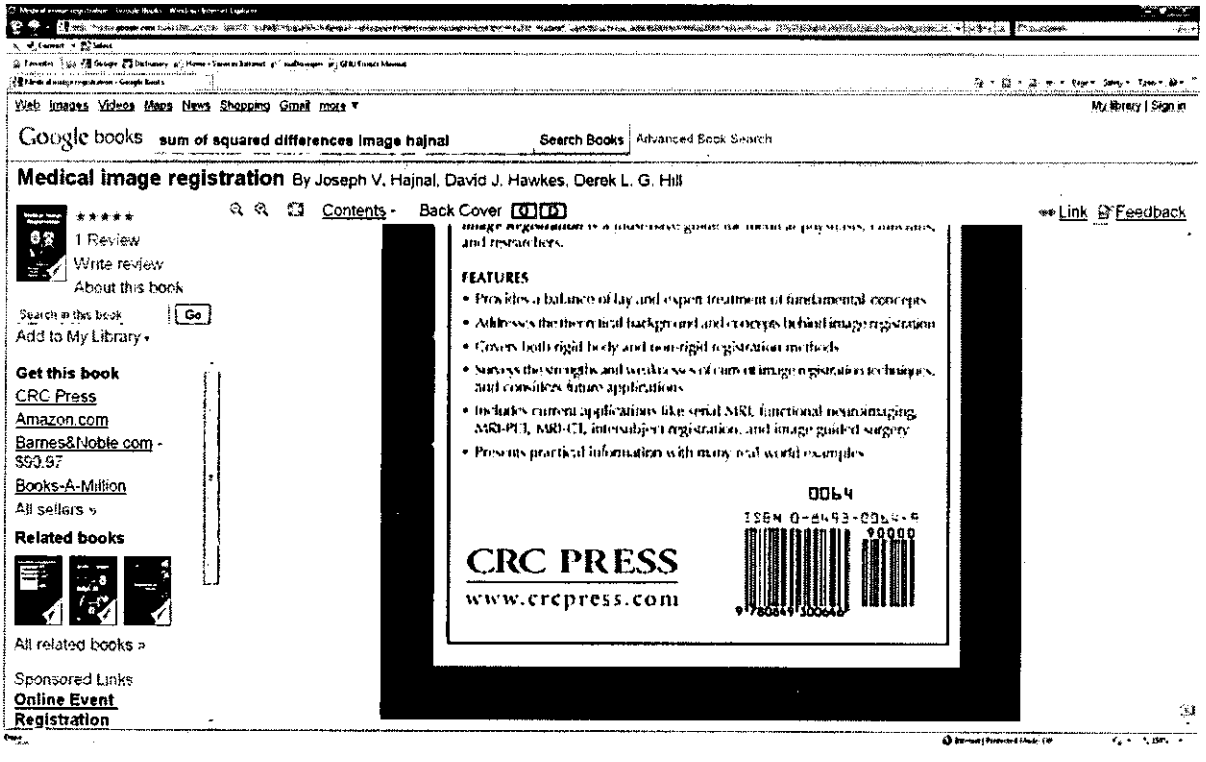
[30] Price RR, Axel L, Morgan T, Newman R, Perman W, Schneiders N, Selikson M, Wood M, Thomas SR, Quality Assurance Methods and Phantoms for Magnetic Resonance Imaging" , AAPM Report No. 28. , Med Phys. 1990 Mar-Apr;17(2):287-95

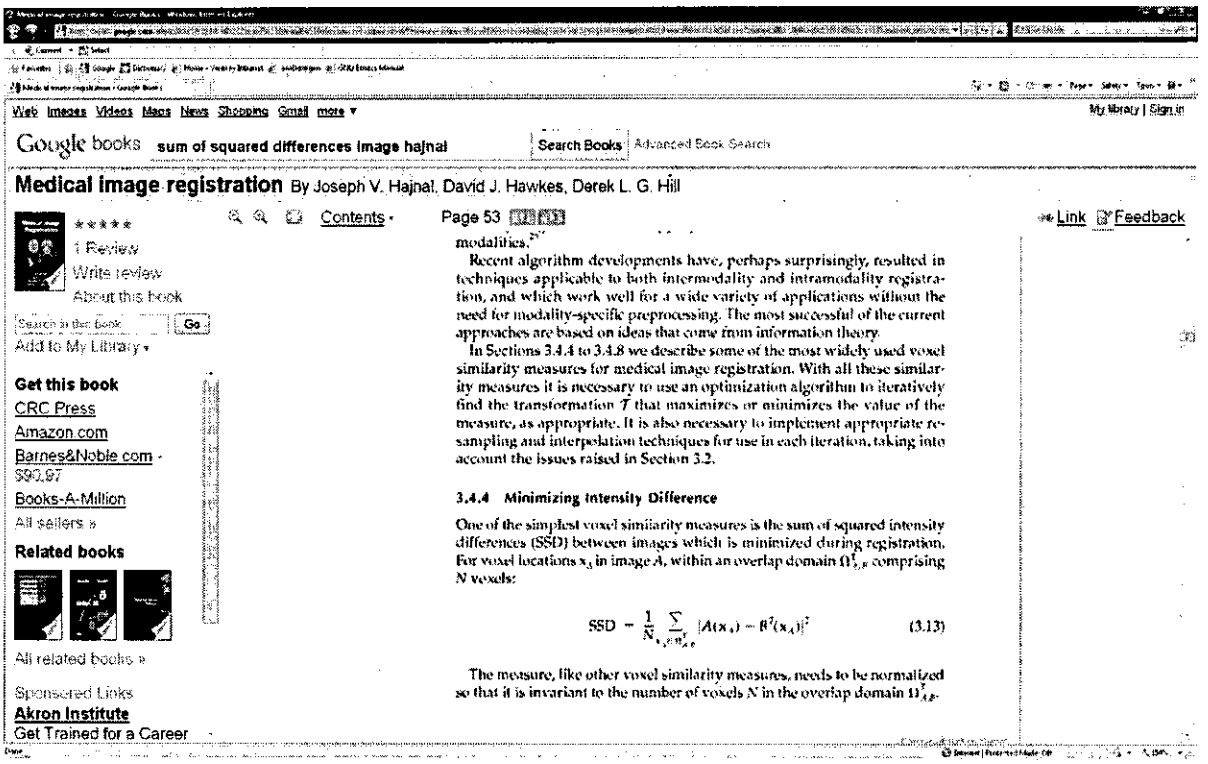
[31] W. Barrett, E. Mortensen, and D.Taylor. An image space algorithm for morphological contour interpolation. In Graphics Interface "94, pages 16-24, 1994

[32] Drzymala RE, Mohan R, Brewster L, Chu J, Goitein M, Harms W, Urie M., Dose-volume histograms, Int J Radiat Oncol Biol Phys. 1991 May 15;21(1):71-8

Medical image registration, J Hajnal, D Hawkes, D Hill, Page 53.







Web Images Videos Maps News Shopping Gmail more ▾ My Library | Sign in

Google books sum of squared differences image hajnal Search Books Advanced Book Search

Medical image registration By Joseph V. Hajnal, David J. Hawkes, Derek L. G. Hill

★★★★★
1 Review
Write review
About this book

Search in the book Go
Add to My Library ▾

Get this book
CRC Press
Amazon.com
Barnes&Noble.com - $90.97
Books-A-Million
All sellers »

Related books

All related books »

Sponsored Links
Akron Institute
Get Trained for a Career

Contents ▾ Page 53 Link Feedback

modalities.[37]

Recent algorithm developments have, perhaps surprisingly, resulted in techniques applicable to both intermodality and intramodality registration, and which work well for a wide variety of applications without the need for modality-specific preprocessing. The most successful of the current approaches are based on ideas that come from information theory.

In Sections 3.4.4 to 3.4.8 we describe some of the most widely used voxel similarity measures for medical image registration. With all these similarity measures it is necessary to use an optimization algorithm to iteratively find the transformation $\hat{T}$ that maximizes or minimizes the value of the measure, as appropriate. It is also necessary to implement appropriate resampling and interpolation techniques for use in each iteration, taking into account the issues raised in Section 3.2.

3.4.4 Minimizing Intensity Difference

One of the simplest voxel similarity measures is the sum of squared intensity differences (SSD) between images which is minimized during registration. For voxel locations $\mathbf{x}_i$ in image A, within an overlap domain $\Omega^{\hat{T}}_{A,B}$ comprising N voxels:

$$\text{SSD} = \frac{1}{N} \sum_{\mathbf{x}_i \in \Omega^{\hat{T}}_{A,B}} [A(\mathbf{x}_i) - B^{\hat{T}}(\mathbf{x}_i)]^2 \qquad (3.13)$$

The measure, like other voxel similarity measures, needs to be normalized so that it is invariant to the number of voxels N in the overlap domain $\Omega^{\hat{T}}_{A,B}$.

DIFFUSION WEIGHTED MAGNETIC RESONANCE IMAGE ANALYSIS AND MEDICAL IMAGE REGISTRATION

By

QINGGUO ZENG

A DISSERTATION PRESENTED TO THE GRADUATE SCHOOL
OF THE UNIVERSITY OF FLORIDA IN PARTIAL FULFILLMENT
OF THE REQUIREMENTS FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY

UNIVERSITY OF FLORIDA

2008

To my family

ACKNOWLEDGMENTS

I would first and foremost like to express my deepest gratitude to my advisor professor Yunmei Chen, for everything she had done for me during my doctoral study. This dissertation would not be possible without her. She provided wonderful research opportunities, invaluable advices on research and life. Dr. Chen introduced me to the field of medical imaging and always helps and encourages me. I have been very lucky to work with her. Her enthusiasm about research stimulates my interests, her insight and experience have guided me through my research. I would also like to thank my other excellent committee members, Dr. James F. Dempsey, Dr. Yijun Liu, Dr. Jay Gopalakrishnan, Dr. Murali Rao for providing numerous advice and discussion. I especially appreciate Dr. James F. Dempsey, Dr. Anneyuko Saito, Dr. Xiaodong Lu and Dr. Jonathan G. Li, Dr. Yijun Liu and Dr. Guojun He for happy cooperations. They provided me lots of invaluable research motivations and clinical data.

I would like to thank Weihong Guo and Feng Huang, for helping me with numerical implementation in my early stage research, without their help, I could not have today's scientific computation ability. I would also like to thank Dr. Xuecheng Tai from University of Bergen, Norway, Dr. Rachid Deriche from INRIA for excellent comments and discussion on my work.

I never would have been able to have finished the dissertation were it not for the unwavering love and support of my family: my wife Weihong, my mum Min Lei, mother in-law Zhongying, brothers Qingjun and Weidong, and sister Shuangxia. Their support and encouragement are my source of strength. I want to especially thank my wife who teaches me the nature of true love. She is an extremely supportive lady, I always appreciate her enthusiasm for life and research. I feel so grateful about my 22-month-old daughter Tienna, who teaches me how wonderful life is by her sunny smiles. I also thank Mr. Shengjing Zhang and Dr. Qing He for their unforgettable guidance before my PhD.

Thank Editorial office for the help I have received in editing this dissertation.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

Abstract of dissertation Presented to the Graduate School
of the University of Florida in Partial Fulfillment of the
Requirements for the Degree of Doctor of Philosophy

DIFFUSION WEIGHTED MAGNETIC RESONANCE IMAGE ANALYSIS AND MEDICAL IMAGE REGISTRATION

By

Qingguo Zeng

May 2008

Chair: Yunmei Chen
Major: Mathematics

My PhD study covers two major topics, one concentrates on Diffusion Weighted Magnetic Resonance Image(DW-MRI) analysis, and the other focuses on medical image registration.

MRI renders non-invasive in vivo information about how water diffuses into a 3D intricate representation of tissues. My work provided histological and anatomical information about tissue structure, composition, architecture, and organization. I have proposed several models to reconstruct human brain white matter fiber tracts, to recover intra-voxel structure, to classify intra-voxel diffusion, to estimate, smooth and characterize apparent diffusion coefficient profiles, and to reconstruct white matter fiber traces.

Medical image registration plays an important role in diagnosis, surgical planning, navigation, and various medical evaluations. But medical images are generally massive and expensive to process, frequently corrupted by high levels of noise, and models to process these data are usually sensitive to choice of parameters. To conquer these challenges, I presented one method to align large volume images efficiently based on a negative log likelihood dissimilarity measure, this method is more robust to the presence of noise and is less sensitive to the choice of weighting parameters. Moreover, this measure is further used in a novel inverse consistent deformable image registration model. Both models are applied on automatic re-contouring on Computed Tomography(CT) images and MRI, and numerical results showed the effectiveness of proposed approaches.

CHAPTER 1
INTRODUCTION

New computer technologies are changing the way medical images is handled and processed, they also make medical image processing more and more involved in research and clinical applications. My study focus on two topics of medical image processing: Diffusion Weighted Magnetic Resonance Image (DW-MRI, also shorten as DWI) analysis and Medical Image Registration.

Chapters 2,3,4 and 5 will cover my study on DW-MRI. In chapters 2 and 3, I will introduce approaches on approximating Apparent Diffusion Coefficients(ADC) by using Spherical Harmonic Series. A method to estimate the probability of water molecule diffusion based on Bi-Gaussian assumption will be introduced in chapter 4, based on which, a fiber trace reconstruction approach is provided in chapter 5. Chapters 6 and 7 cover my work on medical image registration. Future work will be discussed in chapter 8.

In the remaining of the first chapter, overall introductions for all the topics covered in this dissertation will be given, followed by the major contributions for each of chapter 2 through chapter 7.

1.1 ADC Approximation, Diffusion Anisotropy And Diffusion Direction Characterization In Diffusion Weighted Image(DWI)

Diffusion-weighted MRI adds to conventional MRI the capability of measuring the random motion of water molecules in tissue, referred as diffusion. The motion of water molecules can be free or restricted depending on the tissue structures. In tissues containing a large number of fibers (such as cardiac muscle and brain white matter) water diffusion is fastest along the direction that a fiber is pointing to, but slowest in the direction perpendicular to it. This characteristic of the diffusion is termed as anisotropy. In tissues that contain few fibers water diffuses isotropically. DWI renders non-invasively such complex in vivo information about how water diffuses into a 3D intricate representation of tissues, and provides profound histological and anatomical information about tissue structure, composition, architecture, and organization. Changes

in these tissue properties can often be correlated with processes that occur in development, degeneration, disease, and aging, so this technique has become more and more widely applied ([1–4]).

The diffusion of water molecules in tissues over a time interval t can be described by a probability density function $p_t(\mathbf{r})$ on displacement $\mathbf{r}$. Since $p_t(\mathbf{r})$ is largest in the directions of least hindrance to diffusion and smaller in other directions, the information about $p_t(\mathbf{r})$ reveals fiber orientations and leads to meaningful inferences about the microstructure of tissues.

The density function $p_t(\mathbf{r}$ is related to DWI echo signal $s(\mathbf{q})$ via a Fourier transformation (FT) with respect to $\mathbf{q}$, which represents diffusion sensitizing gradient, by

$$s(\mathbf{q}) = s_0 \int p_t(\mathbf{r}) e^{-i\mathbf{q}\cdot\mathbf{r}} d\mathbf{r}, \tag{1–1}$$

where s_0 is MRI signal in the absence of any gradient. Therefore, $p_t(\mathbf{r})$ can be estimated from the inverse FT of $s(\mathbf{q})/s_0$. Recently, Tuch et al. [5] introduced the method of higher angular resolution diffusion(HARD) MRI, and Wedeen et al. [6] succeed in acquiring 500 measurements of $s(\mathbf{q})$ in each scan to perform a fast FT inversion. However, this method requires a large number of measurements of $s(\mathbf{q})$ over a wide range of $\mathbf{q}$ in order to perform a stable inverse FT.

A more common approach to estimate $p_t(\mathbf{r})$ from much sparser set of measurements $s(\mathbf{q})$ is assuming $p_t(\mathbf{r})$ to be a Gaussian. For Gaussian diffusion,

$$p(\mathbf{r}, t) = \frac{1}{\sqrt{(4\pi t)^3 det(D)}} exp\{\frac{-\mathbf{r}^T D^{-1} \mathbf{r}}{4t}\},$$

where D is called the diffusion tensor. Inserting this to equation (1–1) it yields

$$s(\mathbf{q}) = s_0 e^{-b\mathbf{u}^T D\mathbf{u}}, \tag{1–2}$$

where $\mathbf{u} = \mathbf{q}/|\mathbf{q}|$. In this case

$$d(\mathbf{u}) = \mathbf{u}^T D\mathbf{u}.$$

The principle eigenvector of D indicates diffusion direction of the diffusion. The fractional anisotropy (FA) defined as

$$FA = \sqrt{\frac{3}{2}}\sqrt{\frac{(\lambda_1 - \lambda_2)^2 + (\lambda_2 - \lambda_3)^+(\lambda_3 - \lambda_1)^2}{(\lambda_1 + \lambda_2 + \lambda_3)^2}}, \tag{1–3}$$

where λ_i $(i = 1, 2, 3)$ are the eigenvalues of D, has become the most widely used measure of diffusion anisotropy [4]. This is known as diffusion tensor imaging (DTI), and in particular useful for creating white matter fiber tracts [7–11].

However, it has been recognized that the single Gaussian model is inappropriate for assessing multiple fiber tract orientations, when complex tissue structure is found within a voxel [6, 8, 12–16]. A simple extension to non-Gaussian diffusion is to assume that the multiple compartments within a voxel are in slow exchange and the diffusion within each compartment is a Gaussian [13, 14, 17–19]. Under these assumption the diffusion can be modelled by a mixture of n Gaussians:

$$p_t(\mathbf{r}) = \sum_{i=1}^{n} f_i((4\pi t)^3 det(D_i))^{-1/2} e^{\frac{-\mathbf{r}^T D_i^{-1}\mathbf{r}}{4t}}, \tag{1–4}$$

where f_i is the volume fraction of the voxel with the diffusion tensor D_i, $f_i \geq 0$, $\sum_i f_i = 1$, and t is the diffusion time. Inserting (1–4) into equation (1–1) yields

$$s(\mathbf{q}) = s_0 \sum_{i=1}^{n} f_i e^{-b\mathbf{u}^T D_i \mathbf{u}}, \tag{1–5}$$

where $\mathbf{u} = \mathbf{q}/|\mathbf{q}|$, and $b = \gamma^2\delta^2|q|^2(\Delta - \delta/3)$. Here γ is the gyromagnetic ratio, and δ is the duration of two magnetic field gradient pulses with a separation time Δ in the use of Stejskal-Tanner pulsed gradient spin echo method [20]. To estimate D_i and f_i, at least $7n - 1$ measurements $s(\mathbf{q})$ plus s_0 are required. In [17–19] the model of a mixture of two Gaussians were used to estimate the PDF. This estimation requires at least 13 diffusion weighted images from 13 different directions.

One of the alternatives to estimate $p_t(\mathbf{r})$ and characterize diffusion anisotropy is using apparent diffusion coefficient(ADC) profiles $d(\mathbf{x}, \theta, \phi)$, which are related to observed DWI

signals through the Stejskal-Tanner equation:

$$s(\mathbf{q}) = s_0 e^{-bd(\theta,\phi)}, \tag{1–6}$$

where (θ, ϕ) $(0 \leq \theta < \pi, 0 \leq \phi < 2\pi)$ represents the direction of $\mathbf{q}$, b is the diffusion-weighting factor. For Gaussian diffusion $d(\mathbf{u}) = b\mathbf{u}^T D\mathbf{u}$, where $\mathbf{u}$ is the normalized $\mathbf{q}$. The trace, eigenvalues and eigenfunctions of D can be used to characterize the anisotropy and directional properties of the diffusion. For non-Gaussian diffusion the spherical harmonic approximation of the ADC profiles estimated from HARD data has been used for characterization of diffusion anisotropy. This technique was first introduced by Frank [15], also studied by Alexander et al. [21].In the work of [15, 21] $d(\mathbf{x}, \theta, \phi)$ was computed from HARD raw data via the linearized version of (1–6):

$$d(\mathbf{q}) = -\frac{1}{b} log \frac{s(\mathbf{q})}{s_0} \tag{1–7}$$

and represented by a truncated SHS:

$$d(\mathbf{x}, \theta, \phi) = \sum_{l=0,2,l_{max}} \sum_{m=-l}^{l} A_{l,m}(\mathbf{x}) Y_{l,m}(\theta, \phi), \tag{1–8}$$

where $Y_{l,m}(\theta, \phi) : S^2 \rightarrow C$ are the spherical harmonics and C denotes the set of complex numbers. The odd-order terms in the SHS are set to be zero, since the HARD measurements are made by a series of 3-d rotation, $d(\theta, \phi)$ is real and antipodal symmetry. Then, the coefficients $A_{l,m}(\mathbf{x})$'s were used to characterize the diffusion anisotropy. In their algorithm, basically, the voxels with the significant 4th order ($l = 4$) components in SHS are characterized as anisotropic with two-fiber orientations (shorten as two-fibers), while voxels with the significant 2nd order ($l = 2$) but not the 4th order components are classified as anisotropic with single fiber orientation (shorten as one-fiber), which is equivalent to the DTI model. Voxels with the significant 0th order (l=0) but not the 2nd and 4th order components are classified as isotropic. The truncated order is getting higher as the structure complexity increases. Their experimental results showed that

non-Gaussian profiles arise consistently in various brain regions where complex tissue structure is known to exist.

Since the ADC profiles can be used to characterize the diffusion anisotropy, and to estimate $p_t(\mathbf{r})$ through the combination of (1–1) and (1–6), it is of great significance to develop models for better estimation of the ADC profiles from DW MR measurements. In general the raw HARD MRI data are noisy. Computing the coefficients directly from the raw data often provides poor estimates. As a result, it will lead to inaccurate or false characterization of the diffusion and consequently lead to incorrect fiber tracking. The aim of this chapter is to present a novel variational framework for simultaneous smoothing and estimation of non-Gaussian ADC profiles from HARD MRI.

There is growing interest in diffusion tensor denoising and reconstruction from DTI data. There were several popular approaches: (1). Smoothing the raw data $s(\mathbf{q})$ then estimating the diffusion tensor from the smoothed raw data ([22–24]); (2). Smoothing the principal diffusion direction after the diffusion tensor has been estimated from the raw noisy measurements ([25–29]); (3). Smoothing tensor-valued data with preserving the positive definite property of D [23, 27, 30–32].

However, very few research reported in literature to date on HARD data analysis considered denoising problem in the reconstruction of the ADC profiles while the HARD raw data is noisy. To improve the accuracy of the estimation, in chapter 2, we present a novel model that has the ability of simultaneously smoothing and estimating the ADC profile $d(\mathbf{x}, \theta, \phi)$ from the noisy HARD measurements $s(\mathbf{x}, \theta, \phi)$ with preserving the relevant features, and the positiveness and antipodal symmetry constraints of $d(\mathbf{x}, \theta, \phi)$. The basic idea of our approach is to approximate the ADC profiles at each voxel by a 4th order SHS($l_{max} = 4$ in (1–8):

$$d(\mathbf{x}, \theta, \phi) = \sum_{l=0,2,4} \sum_{m=-l}^{l} A_{l,m}(\mathbf{x}) Y_{l,m}(\theta, \phi) \tag{1–9}$$

whose coefficients are determined by solving a constrained minimization problem. This minimization problem minimizes a non-standard growth functional to perform a feature preserved regularization, while it minimizes the data fidelity term. Notice, there are 15 unknown complex valued functions $A_{l,m}$ involved. Since $d(\theta,\phi)$ is real and $Y_{l,m}$ satisfies $Y_{l,-m} = (-1)^m\overline{Y_{l,m}}$, each complex valued $A_{l,m}$ is constrained by $A_{l,-m} = (-1)^m\overline{A_{l,m}}$, where $\overline{A_{l,m}}$ denotes the complex conjugate of $A_{l,m}$. This constraint transforms the 15 unknown complex valued functions in (1–9) to 15 real valued functions: $A_{l,0}(\mathbf{x}), (l = 0, 2, 4), ReA_{l,m}(\mathbf{x}), ImA_{l,m}(\mathbf{x}), (l = 2, 4; m = 1, \ldots, l)$.

Chapter 2 elaborates our work published at [33, 34]. This method differs from the existing approaches developed in [15] and [21] mainly in the aspect of the determination of the $A_{l,m}(\mathbf{x})$'s in (1–8). In ([15]) the $A_{l,m}(\mathbf{x})$'s (l is even) are determined by

$$A_{l,m}(\mathbf{x}) = \int_0^{2\pi}\int_0^{\pi} -\frac{1}{b} log\frac{s(\mathbf{q})}{s_0} Y_{l,m}(\theta,\phi) sin\theta d\theta d\phi, \tag{1–10}$$

and in [21] they are estimated as the least-squares solutions of

$$-\frac{1}{b} log\frac{s(\mathbf{q})}{s_0} = \sum_{l=0}^{l_{max}}\sum_{m=-l}^{l} A_{l,m}Y_{l,m}(\theta,\phi). \tag{1–11}$$

In chapter 2, the estimation of $A_{l,m}(\mathbf{x})$'s is not performed individually at each isolated voxel, but a process of joint estimation and regularization across the entire volume. The joint estimation and regularization not only guarantees the wellposedness of the proposed model, but also enhances the accuracy of the estimation since the HARD data are noisy. Moreover, in chapter 2, we provide more detailed method to characterize the diffusion anisotropy, which uses not only the information of $A_{l,m}(\mathbf{x})$'s as in ([15, 21]), but also the variation of $d(\theta,\phi)$ about its mean. Our experimental results showed the effectiveness of the model in the estimation and enhancement of anisotropy of the ADC profile. The characterization of the diffusion anisotropy based on the reconstructed ADC profiles using the proposed model is consistent with the known fiber anatomy. Detail will be provided in chapter 2.

But, to use SHS model (1–9) to approximate d in (1–6), and hence to detect two-fiber diffusion, at least 15 diffusion weighted measurements $s(\mathbf{q})$ over 15 carefully selected directions are required. However, to use the mixture model (1–4) with n=2 to detect two-fiber diffusion only 13 unknown functions: f, 6 entries of each of D_1, D_2 need to be solved. This motivated us to study what is the minimum number of the diffusion weighted measurements required for detecting diffusion with no more than two fiber orientations within a voxel, and what is the corresponding model to approximate the ADC profiles in this case. In chapter 3(a work published in [35]) we propose to approximate the ADC profiles from HARD MRI by the product of two up to the second order spherical SHS instead of a SHS up to order four. We also show that the product of two up to the second order spherical SHS describes only the diffusion with at most two fiber orientations, while the SHS up to order four may also reveals the diffusion with three fiber orientations.

Moreover, in chapter 3(related work was published in [35]), we will introduce an information measurement developed in [36], and termed as Cumulative Residual Entropy (CRE) (see definition (3–7)) to characterize the diffusion anisotropy. CRE differs from Shannon entropy in the aspect that Shannon entropy depends only on the probability of the event, while CRE depends also on the magnitude of the change of the random variable. We observed that isotropic diffusion has either no local minimum or many local minima with very small variation in the denoised $s(\mathbf{q})/s_0$, i.e., e^{-bd} profiles in comparing with one fiber or two-fiber diffusions, which implies the corresponding CRE to be small. We also found that one fiber diffusion has only one local minimum with larger variation in the $s(\mathbf{q})/s_0$ profiles, which leads to larger CRE. Therefore, we propose to properly threshold the CRE for the regularized $s(\mathbf{q})/s_0$ profiles to characterize the diffusion anisotropy. Detail will be provided in chapter 3.

In chapter 4(related work published in [19]) we present a new variational method for recovering the intra-voxel structure under the assumption that $p_t(\mathbf{r})$ is a mixture of two Gaussians. Our approach differs from the existing methods in the following aspects. First,

we recover each *field* $D_i(\mathbf{x})$ or $f_i(\mathbf{x})$ globally by simultaneous smoothing and data fitting, rather than estimating them from (1–5) with $n = 2$ in each isolated voxel, which leads to an ill-posed problem. Second, we recover the ADC profile $d(\mathbf{x}, \theta, \phi)$ in SH representation using method introduced in chapter 2 from the noisy HARD data before estimating $D_i(\mathbf{x})$ and $f_i(\mathbf{x})$. The recovered d and the voxel classification on diffusion anisotropy from d are incorporated into our energy function to enhance the accuracy of the estimations. Third, we applied the biGaussian model to all the voxels in the field, rather than the voxels where the Gaussian model only fits poorly. Since both the constraint of $f_1 \approx 1$ on the region of strong one-fiber diffusion, and the regularization for f_i and D_i are built in the model, the single fiber and multi-fiber diffusions can be separated automatically by the model solution. This approach should be less sensitive to the error in the voxel classification. See details in chapter 4, related work was published in [19, 37].

1.2 Neuron Fiber Tracts Reconstruction Based On Smooth Tensor Field

The assessment of connectivity and the reconstruction of 3D curves representing fiber traces are useful for basic neuroanatomical research and for disease detection. Most normal brain functions require that specific cortical regions communicate with each other through fiber pathways. Diffusion imaging is based on magnetic resonance imaging technique which was introduced in the mid 1980s [38]and provides a very sensitive probe for detecting biological tissues architecture. The key concept that is of primary importance for diffusion imaging is that diffusion in biological tissues reflects their structure and their architecture at a microscopic scale. For instance, brownian motion is highly influenced in tissues, such as cerebral white matter or the annulus fibrosus of inter-vertebral discs. Measuring, at each voxel, that very same motion along a number of sampling directions(at least six, up to several hundreds) provides an exquisite insight into the local orientation of fibers.

We will assume p_t is a mixture of 2 Gaussians($n = 2$ in (1–4)):

$$s(\mathbf{u}) = s_0(fe^{-b\mathbf{u}^T D_1 \mathbf{u}} + (1-f)e^{-b\mathbf{u}^T D_2 \mathbf{u}}), \tag{1–12}$$

where $f \geq 0, 1-f \geq 0$ are considered as the apparent volume fractions of diffusion tensor D_1, D_2 respectively. Recently, Parker et al [17] and Tuch et al. [18] used a mixture of two Gaussian densities to model the diffusion for the voxels where the Gaussian model fits the data poorly.

A prime problem in recovering multi tensor field $D_i(\mathbf{x}), i = 1, 2$ and $f(\mathbf{x})$ is the acquisition noise which corrupts the data measurement. Neither Parker nor Tuch considered removing noise. In this note we present a new variational method which differs from the existing methods in the following aspects. First, we recover tensor field $D_i(\mathbf{x}), i = 1, 2$ and $f(\mathbf{x})$ globally by simultaneous smoothing and data fitting, rather than estimating them from (1–12) at each isolated voxel, which leads to an ill-posed problem and is impossible to get a smooth multi-tensor field. Second, we applied the biGaussian model to all the voxels in the field while Parker et al [17] and Tuch et al. [18] only applied biGaussian model to the voxels where the Gaussian model fits the data badly. So they need to do preprocessing to distinguish voxels at which Gaussian model fits the data poorly from those at which Gaussian model fits data well. In our approach, this kind of voxel classification is not required and thus avoids the errors coming from it. Section 5.1 will explain in details how to recover smooth multi-diffusion tensor field.

Regarding reconstruction of white matter traces, currently, there are several different approaches which can be roughly divided into four categories: (1) line propagation algorithms; (2) surface propagation algorithms; (3) global energy minimization to find the energetically most favorable path between two predetermined pixels; (4) solving a diffusion equation. We will discuss line propagation algorithms in details. Line propagation algorithms use local information for each step of propagation. The main differences among techniques in this class stem from the kind of local information being considered and the way information from neighboring pixels is incorporated.

The main local information used in most classical algorithms for recovering brain connectivity mapping from DTI data is the principle eigenvector(PE) of the diffusion tensor[11, 39, 40].PE successfully determines the fiber direction in cases where there is a single fiber direction in each voxel, and is therefore adequate for reconstructing large trace systems. However, with voxel sizes typical of diffusion acquisitions($10 - 30mm^3$), there is significant partial volume averaging of fiber direction in anatomical regions of both research and clinical interest, such as the association fibers near the cortex. Moreover, image noise will influence the direction of the major eigenvector. And as degree of anisotropy decreases, the uncertainty in the major eigenvector increases, at which tracking may be erroneous. When diffusion tensor is planar shaped, PE even does not make sense. Westin et.al [41] and Lazar et.al[42] used the entire tensor to deflect the estimated fiber trajectory. This algorithm is called tensor deflection(TEND). The deflection term is better than PE in the sense that the previous one is less sensitive to image noise and is less erroneous in situation of degenerated anisotropy. But it still has the problem of partial volume averaging of fiber direction.

In chapter 5 section 5.2, we will provide a new line propagation algorithm based on smooth multi-tensor field. It keeps all the advantages of TEND and has two additional good properties: first, problem of partial volume averaging is automatically solved as it is based on multi-tensor field; second, it uses dynamically adjusted step size to keep total curvature of traces low, to appropriately terminate tracking and to increase algorithm efficiency. Related work was published in [43–45].

1.3 Medical Image Registration

Image registration, a very important subject in computer vision and image processing, has been increasingly used in image guided surgery, functional brain mapping, multi-modality fusion etc. The task of image registration is to find a transformation h that relates points in the source image S to their corresponding points in the target image T. Recent reviews on registration include Hill et. al. [46], Zitova et. al. [47] and Modersitzki [48]. One

may classify the image registration methods into two categories, namely feature based method and direct method. For a comprehensive survey on feature-based schemes, we refer the reader to Maintz et. al. [49]. Direct methods involve estimating the transformation between two given images from the raw data-with no feature extraction etc. One example of direct method is to find a displacement U, rigid or non-rigid between source images S and target image T such that the following energy is minimized:

$$\int_{\Omega} D(F(S(X+U(X))),T(X)) + \lambda R(U(X))dX \tag{1–13}$$

where U is a geometric transformation, $D(\cdot)$ is a difference measure, $F(\cdot)$ is an intensity transformation, $R(U(X))$ is a regularity term on T and λ is a balancing factor between similarity and smoothness terms. In real applications, how to balance these two terms is quite challenging [48].

For intra-modality cases, D could simply be chosen as L^2 norm of intensity difference, which is also usually called Sum of Squared Differences (SSD). For inter-modality images, a general direct measure is negative Mutual Information(MI)(e.g. [50–53]), one can also transform I_1 and I_2 into same domain, for instance, apply Fourier transform to move both images into the frequency domain, then define D to be L^2 norm of frequency difference. Other transforms, such as Hilbert [54], Gabor could be applied based on specific applications.

Deformable image registration allows more freedom at each point, it has been the subject of extensive study in the literature (e.g. [47, 55–63]). Lu. et. al. [58] present a technique dealing with fast free-form deformation between two images of same modality. In this work, L^2 norms are used both in difference and regularity terms. By using calculus of variations, the minimization problem is reformed into a set of nonlinear elliptic partial differential equations(PDEs). A Gauss-Seidel finite difference scheme is used to solve them iteratively. Vemuri et. al. [64] present a level-set curve evolution technique that lets its

level-sets move along their respective normals with a speed that is proportional to the difference between the target and the evolving source image.

Another method which is reported by Ayache et. al. [65] uses the Demons algorithm of Thirion[56] for achieving non-rigid registration in multi-modal data sets, and techniques for estimating intensity and geometric transformations are provided. Other direct methods could be found in Bajscy et. al. [66], Cachier et. al. [67] and Hermosillo et. al. [52].

In chapter 6, deformable image registration for both same-modality and multi-modality cases will be present in details.

In certain applications such as imaging guided radiation therapy, it would be better to have a one-to-one and inverse consistent deformation field, while the majority of non-rigid registration methods do not guarantee such property. The inverse consistent means that when the source and target images are switched in the model, the point correspondence between S and T does not change. An inverse inconsistent deformation field can generate large errors in the processes like auto re-contouring([68]), dose calculate ([69–73]) in radiation therapy. A number of work have attempted to make the registration inverse consistent (e.g. [74–81]). In this chapter, I will only discuss two of them which are closely related to my work. In [75], Christensen and Johnson proposed the following coupled minimization problems:

$$\begin{aligned} E(h) &= \underbrace{M(S(h),T) + \lambda R(h)}_{E_1} + \rho \int_\Omega |h - g^{-1}|^2 dx \\ E(g) &= \underbrace{M(S,T(g)) + \lambda R(g)}_{E_2} + \rho \int_\Omega |g - h^{-1}|^2 dx \end{aligned} \tag{1–14}$$

where $M(\cdot,\cdot)$ is a dissimilarity measure between two images, g is the backward mapping which deforms T such that $T(g)$ is close to S under measure M, and g is expected to be the inverse of the forward mapping h (i.e. $h \circ g = g \circ h = id$, where id is the identity mapping). $R(\cdot)$ is a regularity measure on deformation fields h and g, $\lambda > 0$ and $\rho > 0$ are parameters balances the goodness of alignment, the smoothness of the deformation,

and consistence of invertibility. g^{-1} and h^{-1} represent numerical inverses of g and h respectively. In [75], h and g were solved by using gradient descent algorithm. In [79], the energies in (1–14) are added together to make the energy function symmetric.

Since the inverse consistent constraints are accommodated by penalty terms in the energy functionals, solutions h and g by (1–14) are not exactly inverse to each other. How h and g are closely inverse to each other depends on how large the parameter ρ is, which in practice is hard to choose and needs to be adjusted case by case. Theoretically, h and g are exactly inverse to each other only when $\rho \to +\infty$.

In [76], Leow et al. proposed a different approach. They find h and g by the time marching scheme:

$$h^{(n+1)} = h^{(n)} + dt(\eta_1 + \eta_2), \quad g^{(n+1)} = g^{(n)} + dt(\xi_1 + \xi_2), \tag{1–15}$$

where dt is the time step, η_1 and ξ_2 are vector fields representing gradient descent directions of E_1 and E_2 in (1) respectively, i.e.

$$\begin{aligned} \eta_1(x) &= -\nabla_h M(S(h(x)), T(x)) - \lambda \nabla_h R(h(x)) \\ \xi_2(x) &= -\nabla_g M(S(x), T(g(x))) - \lambda \nabla_g R(g(x)). \end{aligned} \tag{1–16}$$

To make the model inverse consistent, η_1 and ξ_2 are chosen by the following approach.

Suppose $h^{(n)} \circ g^{(n)} = id$ in the nth iteration, then η_2, ξ_1 were determined by taking care of the inverse consistent constraints $h^{(n+1)} \circ g^{(n+1)} = id$, i.e.

$$(h^{(n)} + dt(\eta_1 + \eta_2)) \circ (g^{(n)} + dt(\xi_1 + \xi_2)) = id. \tag{1–17}$$

Taking the Taylor's expansion of (1–17) with respect to dt at 0, and collecting up to the first order terms of dt, one gets

$$\eta_2(x) = -D(h(x))\xi_2(h(x)), \; \xi_1(x) = -D(g(x))\eta_1(g(x)) \tag{1–18}$$

where D is the Jacobian matrix operator. Relations in (1–18) make the iterations in (1–15) uni-directional, i.e., updating formula for the forward mapping h does not depend on the backward mapping g, and vise versa.

In this scheme, the driving force for updating h (or g) involves both forward force from E_1 and backward force from E_2, so the scheme aligns two images faster than the models in which the force driving the deformation field depends on E_1 or E_2 only([61]).

However, $h^{(n+1)}$ and $g^{(n+1)}$ by (1–15 to 1–18) are not exactly inverse to each other even $h^{(n)}$ and $g^{(n)}$ are. Since in the derivation of (1–18), the higher order terms in the Taylor expansion of the left hand side of (1–17) have been discarded. This generates truncation errors, which are accumulated and exaggerated during iterations. Started with the identity mapping for both h and g, the solutions h and g from (1–15 to 1–18) are not inverse to each other, as we will show in experimental results.

Regarding the dissimilarity measure $M(\cdot,\cdot)$, a conventionally used one for same modality image registration is SSD, which is sensitive to the presence of noise and outliers (e.g. [46, 54]). Moreover, the fixed parameter λ in (1–14) balancing the smoothness of the deformation field and goodness of the alignment is always difficult to select, and affects the robustness of the model to the choice to this weighting parameter. Small λ results an unstable and discontinuous deformation field, while large λ leads to inaccurate result, and may yield a nonphysical deformation field due to unreasonable restrictions. In chapter 7, for a proposed model, I will replace the SSD dissimilarity measure by a likelihood estimation that is based on the assumption of a Gaussian distribution of the residue image between the deformed image and fixed image.

1.4 Contributions

In this section, contributions for each of chapter 2 through chapter 7 will be addressed briefly.

- Estimation, smoothing and characterization of apparent diffusion coefficient using two different methods

Firstly, in Chapter 2 we present a new variational framework for simultaneous smoothing and estimation of apparent diffusion coefficient (ADC) profiles from HARD MRI. The model approximates the ADC profiles at each voxel by a 4th order spherical harmonic series (SHS). The coefficients in SHS are obtained by solving a constrained minimization problem. The smoothing with feature preserved is achieved by minimizing a variable exponent, linear growth functional, and the data constraint is determined by the original Stejskal-Tanner equation. The antipodal symmetry and positiveness of the ADC are accommodated in the model. We use these coefficients and variance of the ADC profiles from its mean to classify the diffusion in each voxel as isotropic, anisotropic with single fiber orientation, or two fiber orientations. The proposed model has been applied to both simulated data and High Angular Resolution Diffusion-weighted (HARD) MRI human brain data. The experiments demonstrated the effectiveness of our method in estimation and smoothing of ADC profiles and in enhancement of diffusion anisotropy. Further characterization of non-Gaussian diffusion based on the proposed model showed a consistency between our results and known neuroanatomy.

Secondly, in Chapter 3 we present another approximation for the ADC of non-Gaussian water diffusion with at most two fiber orientations within a voxel. The proposed model approximates ADC profiles by product of two spherical harmonic series (SHS) up to order 2 from HARD MRI data. The coefficients of SHS are estimated and regularized simultaneously by solving a constrained minimization problem. An equivalent but non-constrained version of the approach is also provided to reduce the complexity and increase the efficiency in computation. Compare to the first method, this one requires less measurements but provides comparable results. Moreover we use the CRE as a measurement to characterize diffusion anisotropy. By using CRE we can get reasonable results using two thresholds, while the existing methods either only can be used to characterize Gaussian diffusion or need more measurements

and thresholds to classify anisotropic diffusion with two fiber orientations. The experiments on HARD MRI human brain data indicate the effectiveness of the method in the recovery of ADC profiles. The characterization of diffusion based on the proposed method shows a consistency between our results and known neuroanatomy.

- Reconstruction of Intra-voxel Structure from Diffusion Weighted Images

  In chapter 4, we present a new variational method for recovering the intra-voxel structure under the assumption that $p_t(\mathbf{r})$ is a mixture of two Gaussians: $p_t(\mathbf{r}) = \sum_{i=1}^{2} f_i((4\pi t)^3 det(D_i))^{-1/2} e^{\frac{-\mathbf{r}^T D_i^{-1}\mathbf{r}}{4t}}$. Our approach differs from the existing methods in the following aspects. First, we recover each *field* $D_i(\mathbf{x})$ or $f_i(\mathbf{x})$ globally by simultaneous smoothing and data fitting, rather than estimating them from (1–5) with $n = 2$ in each isolated voxel, which leads to an ill-posed problem. Second, we recover the ADC profile $d(\mathbf{x}, \theta, \phi)$ in SH representation using method introduced in chapter 2 from the noisy HARD data before estimating $D_i(\mathbf{x})$ and $f_i(\mathbf{x})$. The recovered d and the voxel classification on diffusion anisotropy from d are incorporated into our energy function to enhance the accuracy of the estimates. Third, we apply the biGaussian model to all the voxels in the field, rather than the voxels where the Gaussian model only fits poorly. Since both the constraint of $f_1 \approx 1$ on the region of strong one-fiber diffusion, and the regularization for f_i and D_i are built in the model, the single fiber and multi-fiber diffusions can be separated automatically by the model solution. This approach should be less sensitive to the error in voxel classification.

- Reconstruct White Matter Fiber Traces Using Multi-Tensor Deflection In DWI

  In chapter 5, we will provide a new line propagation algorithm based on smooth multi-tensor field.

  We assume there are up to two diffusion channels at each voxel. A variational framework for 3D simultaneous smoothing and recovering of multi-diffusion tensor

field as well as a novel multi-tensor deflection(MTEND) algorithm for extracting white matter fiber traces based on multi-tensor field are provided. MTEND keeps all the advantages of TEND and has two additional good properties: first, problem of partial volume averaging is automatically solved as it is based on a multi-tensor field; second, it uses a dynamically adjusted step size to keep total curvature of traces low, to appropriately terminate tracking and to increase algorithm efficiency.

Fiber traces are colored using Laplacian eigenmaps. By applying the proposed model to synthetic data and human brain high angular resolution diffusion magnetic resonance images(MRI) data of several subjects, we show the effectiveness of the model in recovering intra-voxel multi-fiber diffusion and inter-voxel fiber traces. Superiority of the proposed model over existing models are also demonstrated.

- Efficient and Robust deformable image registration

  In chapter 6, a fast, robust and accurate deformable image registration technique is provided. The proposed method considers the difference between the deformed template and the reference at each voxel as independent normal random variables with zero mean and a variance to be optimized. Then the optimal deformation field and variance are obtained by minimizing an energy function that consists of the negative log-likelihood function and a regularization term for the deformation field. Since the proposed model accommodates the variance in the alignment, it is more general and robust to noise than the widely used Sum of Square Distance (SSD) model, which has limitation in registering non-linearly related images. Moreover, the weight between the smoothness of the deformation field and goodness of the alignment in the proposed model is adjusted automatically by the updated variance during iterations. This feature improves the accuracy of the alignment and speed of the convergence. To enhance algorithm efficiency we used the Additive Operator Splitting(AOS), a fast semi-implicit finite difference scheme, and multi-resolution method in solving the minimization problem. The experimental results on synthetic

images, and volumetric CT and MRI lung data indicate the effectiveness of the proposed scheme.

- Inverse consistent deformable image registration and its application on automatic re-contouring

  Chapter 7 provides a novel algorithm for invertible non- rigid image registration. The proposed model minimizes two energy functionals coupled by a natural inverse consistent constraint. Both of the energy functionals for forward and backward deformation fields consist a smoothness measure of the deformation field, and a similarity measure between the deformed image and the one to be matched. In this proposed model the similarity measure is based on maximum likelihood estimation of the residue image. To enhance algorithm efficiency, the Additive Operator Splitting (AOS) scheme is used in solving the minimization problem. The inverse consistent deformation field can be applied to automatic re-contouring to get an accurate delineation of Regions Of Interest(ROIs). The experimental results on synthetic images and 3D prostate data indicate the effectiveness of the proposed method in inverse consistency and automatic re-contouring. Related work will be published in [82].

CHAPTER 2
ESTIMATION,SMOOTHING AND CHARACTERIZATION OF ADC IN HARD WEIGHTED IMAGE USING SPHERICAL HARMONIC SERIES UP TO ORDER 4

2.1 Model Description

We will present a variational framework for simultaneous smoothing and estimation of the ADC profile $d(x, \theta, \phi)$ from the noisy HARD measurements $s(\mathbf{q})$. This model minimizes an energy functional consisting of two parts: one is the cost for regularizing d, and the other is the cost for fitting d to the HARD measurements through the original Stejskal-tanner equation (1–6). To explain the basic idea of our method, we focus our attention on the cases where there are at most two fibers passing through a single voxel. The same idea can be applied to the cases where there are more fibers within a voxel.

The challenge in regularizing d comes from two aspects. First, d is defined on $\Omega \times S^2$ rather than $\Omega \times R^2$, hence, the derivatives for (θ, ϕ) should be along the sphere. Secondly, the regularized d has to preserve the antipodal symmetry property with respect to (θ, ϕ), since the HARD measurements are made by a series of 3-d rotation. Considering these facts we adopt the idea developed in [15, 21] that approximates d by its SHS consisting of only even order components up to order 4 , i.e.

$$d(x, \theta, \phi) = \sum_{l=0,2,4} \sum_{m=-l}^{l} A_{l,m}(\mathbf{x}) Y_{l,m}(\theta, \phi). \tag{2–1}$$

The expression in (2–1) ensures the smoothness and antipodal symmetry property of $d(\mathbf{x}, \theta, \phi)$ in terms of (θ, ϕ), this is easy to see from the definition of $Y_{l,m}(\theta, \phi)$. For the cases where possibly k fibers cross in a single voxel, the sum in (2–1) should be replaced by $\sum_{l=0,2,\ldots,2k}$.

Now the problem of regularization and estimation of $d(x, \theta, \phi)$ reduces to that for the 15 complex valued functions $A_{l,m}(x)$ ($l = 0, 2, 4$ and $m = -l, \ldots, l$) in (2–1). Since $d(\theta, \phi)$ at each voxel is a real valued function, and $Y_{l,m}$ satisfies $Y_{l,-m} = (-1)^m Y^*_{l,m}$, $A_{l,m}$ should be

constrained by

$$A_{l,-m} = (-1)^m A^*_{l,m}.$$

This constraint reduces the number of the unknown coefficients $A_{l,m}$ in (2–1) to 15 real valued functions. They are

$$A_{l,0}(x),\ (l = 0, 2, 4),\quad ReA_{l,m}(x),\quad ImA_{l,m}(x),\quad (l = 2, 4\ and\ m = 1, \ldots, l). \tag{2–2}$$

By using (2–2), we can rewrite (2–1) as

$$d(x, \theta, \phi) = \sum_{l=0,2,4} A_{l,0}(\mathbf{x}) Y_{l,0}(\theta, \phi) + 2 \sum_{l=2,4} \sum_{m=1}^{l} (ReA_{l,m}(x) ReY_{l,m}(\theta, \phi) - ImA_{l,m}(x) ImY_{l,m}(\theta, \phi)), \tag{2–3}$$

where ReF and ImF represent the real and imaginary part of a function F respectively. Now the problem of regularizing and estimating d reduces to smoothing and estimation of 15 functions in (2–2) simultaneously.

There are many choices of regularizing operators to smooth the 15 functions in (2–2). Total Variation (TV) based regularization, first proposed by Rudin, Osher and Fatemi[83], proved to be an invaluable tool for feature preserving smoothing. However, it sometimes causes a staircase effect making restored image blocky, and even containing 'false edges' [84, 85]. An improvement, that combines the TV based smoothing with isotropic smoothing, was given by Chambolle and Lions [85]. Their model minimizes the TV norm when the magnitude of the image gradient is larger, and the L^2 norm of the image gradient if it is smaller. However, this model is sensitive to the choice of the threshold which separates the TV based and isotropic smoothing. To further improve Chambolle and Lions' model and make the model having an ability to self adjust diffusion property, recently, certain nonstandard diffusion models based on minimizing $L^{p(x)}$ norm of image gradient have been developed [84, 86]. To recover an image u from an observed

image I in [84] the diffusion was governed by minimizing

$$\min_u \int_\Omega |\nabla u|^{p(|\nabla u|)}$$

where $p(s)$ is monotonically decreasing function and $\lim_{s\to 0} p(s) = 2$, $\lim_{s\to\infty} p(s) = 1$. In [86] the diffusion was performed through minimizing

$$\min_u \int_\Omega \phi(x, Du)$$

where $\phi(x,r) := \frac{1}{p(x)}|r|^{p(x)}$ if $|r| \le a$, and $\phi(x,r) := |r| - \frac{ap(x)-a^{p(x)}}{p(x)}$, if $|r| > a$ for threshhold a. In this model $p(x)$ is chosen as

$$p(x) = p(|\nabla I|) = 1 + \frac{1}{1 + k|\nabla G_\sigma * I|^2},$$

where $k,\ \sigma > 0$ are parameters, G_σ is the Gaussian kernel. Both models in [84] and [86] are able to self adjust diffusion range from isotropic to TV-based depending on image gradient. At the locations with higher image gradients ($p = 1$), the diffusion is TV based and strictly tangential to the edges ([83, 85, 87]). In homogeneous regions the image gradients are very small ($p = 2$), the diffusion is essentially isotropic. At all other locations, the image gradient forces $1 < p < 2$, and the diffusion is between isotropic and total variation based and varies depending on the local properties of the image. This self adjusting ability enables these models effectively preserving features when images are smoothed.

Applying the idea of minimizing functionals with variable exponent to the problem of regularizing the coefficients $A_{l,m}$'s in (2–1) , we propose to minimize

$$E_1(A_{l,m}) =: \int_\Omega \sum_{l=0,2,4} \sum_{m=-l}^{l} \phi_{l,m}(x, DA_{l,m}), \tag{2–4}$$

where

$$\phi_{l,m}(x,r) := \begin{cases} \frac{1}{p_{l,m}(x)}|\frac{r}{M_{l,m}}|^{p_{l,m}(x)}, & |r| \le M_{l,m} \\ |\frac{r}{M_{l,m}}| - (1 - \frac{1}{p_{l,m}(x)}), & |r| > M_{l,m} \end{cases} \tag{2–5}$$

In (2–5)

$$p_{l,m} = 1 + \frac{1}{1 + k|\nabla G_\sigma * a_{l,m}(x)|^2}, \tag{2–6}$$

and $a_{l,m}$ is the least-squares solution of

$$-\frac{1}{b} log \frac{s(x,\theta,\phi)}{s_0(x)} = \sum_{l=0,2,4} \sum_{m=-l}^{l} a_{l,m}(x) Y_{l,m}(\theta,\phi). \tag{2–7}$$

In real application, $M_{l,m}$ is picked as 90% percentile of r. Experimental results are not sensitive to this parameter 90%.

Again we would like to point out that E_1 only needs to include 15 terms corresponding to the 15 real valued functions in (2–2). Here we write it in terms of $A_{l,m}$ in order to shorten the expression of the formula. If using $A_{l,m}$ instead of $a_{l,m}$ in (2–6) we may get better numerical results, since $p_{l,m}$ would depend on updated $A_{l,m}$ rather than the fixed $a_{l,m}$ in the iterations to minimize (2–4). However, it gives difficulty in the study of the existence of solutions.

Since $d(x,\theta,\phi)$ is related to the HARD measurements $s(x,\theta,\phi)$ and $s_0(x)$ through the Stejskal-tanner equation (1–6), the estimation of the $A_{l,m}$'s is based on the original Stejskal-tanner equation (1–6) rather than its (log) linearized form (1–7), that is,

$$E_2(A_{l,m}) =: \frac{1}{2} \int_\Omega \int_0^{2\pi} \int_0^{\pi} |s(\mathbf{x},\theta,\phi) - s_0(x) e^{-bd(x,\theta,\phi)}|^2 sin\theta d\theta d\phi dx, \tag{2–8}$$

where d is determined in (2–1). As observed in [88] when the signal to noise ratio is low the linearized model gives different results.

Finally, to simultaneously regularize and estimate the ADC $d(x,\theta,\phi)$, our model minimizes the energy function

$$E(A_{l,m}) =: \lambda E_1(A_{l,m}) + E_2(A_{l,m}), \tag{2–9}$$

with respect to $A_{l,m}$ ($l = 0, 2, 4$ and $m = -l, \ldots, l$) in the space of $BV(\Omega)$, (in fact, only 15 functions in (2–2) are needed), and subject to the constraint:

$$d(x, \theta, \phi) \geq 0. \tag{2–10}$$

In (2–9),(2–10), $s(x, \theta, \phi)$ and $s_0(\mathbf{x})$ are the noisy HARD measurements (real valued), $d(x, \theta, \phi)$ is the SHS given in (2–1), $\Omega \subset R^3$ is the image domain, $\lambda > 0$ is a parameter which could be different for different A_{lm}. E_1 and E_2 are given in (2–4) and (2–8), respectively. We would like to point out that for a function $A_{l,m} \in BV$, $DA_{l,m}$ is a measure, the definition of (2–4) is not obvious. This will be discussed in existence section 2.5 below.

Before we derive the Euler-Lagrange equations for our model (2–9),(2–10), we would like to point out that if the measurements satisfy the condition $s(x, \theta, \phi) \leq s_0(x)$, the solution of (2–9) meets the constraint (2–10) automatically. Therefore we can treat our model as an unconstrained minimization. This is given in the following lemma.

Lemma: Under the assumption that

$$s(x, \theta, \phi) \leq s_0(x), \quad for\ all\ \ \mathbf{x} \in \Omega,\ 0 \leq \theta < \pi,\ 0 \leq \phi < 2\pi \tag{2–11}$$

the minimizer of (2–9) always satisfies the constraint (2–10).

Proof: Let $A_{l,m}(x)$ ($l = 0, 2, 4$ and $m = -l, \ldots, l$) be the minimizer of (2–9) in $BV(\Omega)$, and $d(x, \theta, \phi)$ be the function defined in (2–1) associated with these optimal $A_{l,m}(x)$'s. If for some $x \in \Omega,\ 0 \leq \theta < \pi,\ 0 \leq \phi < 2\pi$, $d(\mathbf{x}, \theta, \phi) < 0$, we define $\hat{d}(x, \theta, \phi)$:

$$\hat{d}(x, \theta, \phi) := \begin{cases} d(x, \theta, \phi), & if\ d(x, \theta, \phi) \geq 0 \\ 0, & if\ d(x, \theta, \phi) < 0 \end{cases} \tag{2–12}$$

and

$$\hat{A}_{l,m}(x) =: \int_0^{2\pi} \int_0^{\pi} \hat{d}(x, \theta, \phi) Y_{l,m}(\theta, \phi) sin\theta d\theta d\phi.$$

Then, using the orthonormality of the spherical harmonics and the definition of d, we have

$$\hat{A}_{l,m}(x) = \begin{cases} A_{l,m}(x), & if\ d(x,\theta,\phi) \geq 0 \\ 0, & if\ d(x,\theta,\phi) < 0 \end{cases} \tag{2–13}$$

This implies that

$$\phi_{l,m}(x, D\hat{A}_{l,m}) \leq \phi_{l,m}(x, DA_{l,m}),$$

hence

$$E_1(\hat{A}_{l,m}) \leq E_1(A_{l,m}).$$

Moreover, it is easy to obtain

$$E_2(\hat{d}) \leq E_2(d),$$

if (2–11) holds. From the last two inequality above, we obtain that $E(\hat{d}) \leq E(d)$. This contradicts to the fact that d minimizes energy functional (2–9).

Now we give the evolution equations associated with the Euler-Lagrange (EL) equations for (2–9): for $l = 0, 2, 4$ and $m = -l, \ldots, l$,

$$\frac{\partial A_{l,m}}{\partial t} = \lambda div(\partial_r \phi_{l,m})(x, DA_{l,m}) - b \int_0^{2\pi} \int_0^{\pi} s_0 e^{-bd}(s - s_0 e^{-bd}) Y_{l,m} sin\theta d\theta d\phi, \tag{2–14}$$

with the initial and boundary conditions:

$$A_{l,m} = a_{l,m}, \qquad on\ \ \Omega \times \{t = 0\},$$

$$(\partial_r \phi_{l,m})(x, DA_{l,m}) \cdot n = 0 \quad on\ \ \partial\Omega \times \Re^+.$$

In the above EL equation n is the unit outward normal to the boundary of Ω, and $\phi_r(x, r)$ is a continuously differentiable function in r, and

$$\partial_r \phi_{l,m}(x, r) := \begin{cases} \frac{1}{M_{l,m}^P} |r|^{p(x)-2} r, & |r| \leq M_{l,m} \\ \frac{1}{M_{l,m}} |r|^{-1} r, & |r| > M_{l,m} \end{cases} \tag{2–15}$$

$\partial_r \phi_{l,m}$ can also be written as

$$\partial_r \phi_{l,m}(x,r) := \frac{1}{M_{l,m}^q} |r|^{q(x)-2} r, \tag{2–16}$$

where $q(x) = p(x)$ if $|r| \leq M_{l,m}$, and $q(x) = 1$ if $|r| > M_{l,m}$.

2.2 Characterization Of Anisotropy

In [15] the $|A_{l,m}(x)|$ ($l = 0, 2, 4$ and $m = -l, \ldots, l$) in the truncated SHS (2–1) are used to characterize the diffusion anisotropy at each voxel x. Our experimental results, however, indicate this information alone is insufficient to separate isotropic diffusion, one-fiber diffusion, and multi-fibers diffusion within a voxel. We propose to combine the information from $|A_{l,m}|$ with the variances of $d(\phi, \theta)$ about its mean value to characterize the diffusion anisotropy. We outlined our algorithm as follows:

(1). If

$$R_0 =: |A_{0,0}| / \sum_{l=0,2,4} \sum_{m=-l}^{l} |A_{l,m}|, \tag{2–17}$$

is large, or the variance of $d(\theta, \phi)$ about its mean is small, the diffusion at such voxels is classified as isotropic.

(2). For the remaining voxels, if

$$R_2 =: \sum_{m=-2}^{m=2} |A_{2,m}| / \sum_{l=2,4} \sum_{m=-l}^{l} |A_{l,m}| \tag{2–18}$$

is large, the diffusion at such voxels is characterized as one-fiber diffusion. Figure 2-3d presents an intensity-coded image of R_2 in a brain slice through the external capsule, an important structure of the human white matter. In Figure 2-3d those voxels of a high intensity (bright regions on the image) are characterized as one-fiber diffusion.

(3). For each uncharacterized voxel after the above two steps, search the directions (θ, ϕ), where $d(\theta, \phi)$ attains its local maxima. Then we compute the weights for the local maxima (say we have 3 local maxima):

$$W_i =: \frac{d(\theta_i, \phi_i) - d_{min}}{\sum_{i=1}^{3} d(\theta_i, \phi_i) - 3 d_{min}},$$

where (θ_i, ϕ_i) $(i = 1, 2, 3)$ are the directions in which d attains 3 local maxima. If one of the weights is significant, it is considered as one fiber diffusion. If two weights are similar but much larger than the third one, it is viewed as two-fiber diffusion, if all three weights are similar, d can be considered either three-fibers diffusion or isotropic diffusion. In our experiment we restrict ourselves to the cases where we only distinct isotropic, one-fiber or two-fibers diffusions. Under this restriction if three weights are similar, we include this voxel in the class of isotropic diffusion. Figure 2-5a shows our classification of isotropic diffusion (dark region), one-fiber diffusion (gray region), and two-fibers diffusion (bright region) in the same slice as in Figure 2-3.

2.3 Numerical Implementation Issues

To efficiently solve the Euler-Lagrange equation Equations (2–14), we use Additive Operator Splitting(AOS) algorithm for the diffusion operator (see [89, 90]). By using this algorithm, the computational and storage cost is linear in the number of voxels, and the computational efficiency can be increased by a factor of 10 under realistic accuracy requirements([89]). In this algorithm the processes of solving $A_{l,m}$ in different dimensions are independent from each other in every iteration, the algorithm is ready to be modified to a paralleled version.

To avoid the complication of notations we use X to represent any $A_{l,m}$ in the Euler-Lagrange equations, and only write the algorithm for one of the equations (2–14) in the system, since each equation has the same structure as others.

We use semi-implicit finite difference scheme:

$$\frac{X_{i,j}^{(n+1)} - X_{i,j}^{(n)}}{\tau} = f(X_{i,j}^{(n)}) + \lambda div\left(\frac{\nabla X_{i,j}^{(n+1)}}{M^{q_{ij}}|\nabla X_{i,j}^{(n)}|^{2-q_{ij}}}\right)$$

$$= f(X_{i,j}^{(n)}) + \frac{-\lambda ln M}{M^{q_{ij}}}\frac{\nabla q_{ij} \cdot \nabla X_{i,j}^{(n+1)}}{|\nabla X_{i,j}^{(n)}|^{2-q_{ij}}} + \frac{\lambda}{M^{q_{ij}}} div\left(\frac{\nabla X_{i,j}^{(n+1)}}{|\nabla X_{i,j}^{(n)}|^{2-q_{ij}}}\right) \qquad (2\text{–}19)$$

Here X can be replaced by one of $A_{l,m}$'s with $l = 0, 2, 4,\ m = -l \cdots l$, and f is a function of results from last iteration,namely, f is a function of all $A_{l,m}^{(n)}$'s. $q(x) = p(x)$ if

$|\nabla X| \leq M$, and $q(x) = 1$ if $|\nabla X| > M$ for some fixed constant M, which was chosen based on initial value of X, so M might be different for different $A_{l,m}$'s.

For simplicity of formulas, we define:

$$\triangle^x_- X_{i,j} = X_{i,j} - X_{i-1,j},\ \triangle^x_+ X_{i,j} = X_{i+1,j} - X_{i,j},\ \triangle^x X_{i,j} = X_{i+1,j} - X_{i-1,j}$$

$$\triangle^y_+ X_{i,j} = X_{i,j+1} - X_{i,j},\ \triangle^y_- X_{i,j} = X_{i,j} - X_{i,j-1},\ \triangle^y X_{i,j} = X_{i,j+1} - X_{i,j-1}$$

Adopting a discretization of the divergence operator from [91], one can write (2–19) as:

$$\frac{X^{(n+1)}_{i,j} - X^{(n)}_{i,j}}{\tau} = f(X^{(n)}_{i,j}) - \frac{\lambda ln M}{M^{q_{ij}}} \frac{[\triangle^x q_{ij}, \triangle^y q_{ij}]}{2h} \cdot \frac{\left[\triangle^x X^{(n+1)}_{i,j}, \triangle^y X^{(n+1)}_{i,j}\right]/(2h)}{\left(\frac{(\triangle^x X^{(n)}_{i,j})^2}{(2h)^2} + \frac{(\triangle^y X^{(n)}_{i,j})^2}{(2h)^2}\right)^{\frac{2-q_{ij}}{2}}} + \frac{\lambda}{M^{q_{ij}} h^2} \cdot$$

$$\left[\triangle^x_- \left(\frac{\triangle^x_+ X^{(n+1)}_{i,j}}{\left(\frac{(\triangle^x_+ X^{(n)}_{i,j})^2}{h^2} + \frac{(\triangle^y X^{(n)}_{i,j})^2}{(2h)^2}\right)^{\frac{2-q_{ij}}{2}}}\right) + \triangle^y_- \left(\frac{\triangle^y_+ X^{(n+1)}_{i,j}}{\left(\frac{(\triangle^y_+ X^{(n)}_{i,j})^2}{h^2} + \frac{(\triangle^x X^{(n)}_{i,j})^2}{(2h)^2}\right)^{\frac{2-q_{ij}}{2}}}\right)\right]$$

$$= f(X^{(n)}_{i,j}) + (C_{i,j} - G_{i,j}) X^{(n+1)}_{i-1,j} - (C_{i,j} + D_{i,j}) X^{(n+1)}_{i,j} + (D_{i,j} + G_{i,j}) X^{(n+1)}_{i+1,j} +$$

$$+ \ (E_{i,j} - H_{i,j}) X^{(n+1)}_{i,j-1} - (E_{i,j} + F_{i,j}) X^{(n+1)}_{i,j} + (F_{i,j} + H_{i,j}) X^{(n+1)}_{i,j+1} \qquad (2\text{–}20)$$

Where C, D, E and F are from divergence operation, while G and H are generated by dot product, in detail:

$$C_{i,j} = \frac{\lambda}{M^{q_{ij}} h^2} \left[\frac{(X^{(n)}_{i,j} - X^{(n)}_{i-1,j})^2}{h^2} + \frac{(X^{(n)}_{i-1,j+1} - X^{(n)}_{i-1,j-1})^2}{(2h)^2}\right]^{\frac{q_{i-1,j}-2}{2}}$$

$$D_{i,j} = \frac{\lambda}{M^{q_{ij}} h^2} \left[\frac{(X^{(n)}_{i+1,j} - X^{(n)}_{i,j})^2}{h^2} + \frac{(X^{(n)}_{i,j+1} - X^{(n)}_{i,j-1})^2}{(2h)^2}\right]^{\frac{q_{ij}-2}{2}}$$

$$E_{i,j} = \frac{\lambda}{M^{q_{ij}} h^2} \left[\frac{(X^{(n)}_{i+1,j-1} - X^{(n)}_{i-1,j-1})^2}{(2h)^2} + \frac{(X^{(n)}_{i,j} - X^{(n)}_{i,j-1})^2}{(h)^2}\right]^{\frac{q_{i,j-1}-2}{2}}$$

$$F_{i,j} = \frac{\lambda}{M^{q_{ij}} h^2} \left[\frac{(X^{(n)}_{i+1,j} - X^{(n)}_{i-1,j})^2}{(2h)^2} + \frac{(X^{(n)}_{i,j+1} - X^{(n)}_{i,j})^2}{(h)^2}\right]^{\frac{q_{ij}-2}{2}}$$

$$G_{i,j} = -\frac{\lambda\, lnM(q_{i+1,j}-q_{i-1,j})}{M^{q_{ij}}(2h)^2}\left[\frac{(X^{(n)}_{i+1,j}-X^{(n)}_{i-1,j})^2}{(2h)^2}+\frac{(X^{(n)}_{i,j+1}-X^{(n)}_{i,j-1})^2}{(2h)^2}\right]^{\frac{q_{ij}-2}{2}}$$

$$H_{i,j} = -\frac{\lambda\, lnM(q_{i,j+1}-q_{i,j-1})}{(2h)^2}\left[\frac{(X^{(n)}_{i+1,j}-X^{(n)}_{i-1,j})^2}{(2h)^2}+\frac{(X^{(n)}_{i,j+1}-X^{(n)}_{i,j-1})^2}{(2h)^2}\right]^{\frac{q_{ij}-2}{2}}$$

Solving (2–20) would involve matrix inverse operation, which will become more and more complicated and dramatically expensive as dimension increases if we solve it directly. Instead, here we use Additive Operator Splitting (AOS) algorithm, then system (2–20) can be reformatted into the following two systems for the first and second dimensions of X respectively:

$$\frac{\bar{X}^{(n+1)}_{i,j}-X^{(n)}_{i,j}}{\tau} = f(X^{(n)}_{i,j})+$$
$$2\left[(C_{i,j}-G_{i,j})\bar{X}^{(n+1)}_{i-1,j}-(C_{i,j}+D_{i,j})\bar{X}^{(n+1)}_{i,j}+(D_{i,j}+G_{i,j})\bar{X}^{(n+1)}_{i+1,j}\right] \quad (2\text{–}21)$$

$$\frac{\bar{\bar{X}}^{(n+1)}_{i,j}-X^{(n)}_{i,j}}{\tau} = f(X^{(n)}_{i,j})+$$
$$2\left[(E_{i,j}-H_{i,j})\bar{\bar{X}}^{(n+1)}_{i,j-1}-(E_{i,j}+F_{i,j})\bar{\bar{X}}^{(n+1)}_{i,j}+(F_{i,j}+H_{i,j})\bar{\bar{X}}^{(n+1)}_{i,j+1}\right] \quad (2\text{–}22)$$

and

$$X^{(n+1)}_{i,j} = \frac{\bar{X}^{(n+1)}_{i,j}+\bar{\bar{X}}^{(n+1)}_{i,j}}{2}$$

To accommodate the boundary condition $\frac{\partial X}{\partial n}=0$ for the $M\times N$ matrix X, one needs to have:

$$X^{(n+1)}_{1,j} = X^{(n+1)}_{2,j}, X^{(n+1)}_{M-1,j} = X^{(n+1)}_{M,j}$$
$$X^{(n+1)}_{i,1} = X^{(n+1)}_{i,2}, X^{(n+1)}_{i,N-1} = X^{(n+1)}_{i,N}$$

then (2–21) and (2–22) correspond to linear systems in matrix-vector notation:

$$A_1\underline{\bar{X}}^{(n+1)} = \tau f(\underline{\bar{X}}^{(n)})+\underline{\bar{X}}^{(n)}$$

$$A_2\underline{\bar{\bar{X}}}^{(n+1)} = \tau f(\underline{\bar{\bar{X}}}^{(n)})+\underline{\bar{\bar{X}}}^{(n)}$$

where $\underline{\bar{X}}$ and $\underline{\bar{\bar{X}}}$ are $(M-2)(N-2) \times 1$ vectors formed by columns and transpose of rows of the original matrix X respectively, both A_1 and A_2 are $(M-2)(N-2) \times (M-2)(N-2)$ matrices, specifically, A_1 is a tri-diagonal matrix that repeats a $(M-2)\times (M-2)$ tri-diagonal matrix $(N-2)^2$ times diagonally, and A_2 is a tri-diagonal matrix that repeats a $(N-2) \times (N-2)$ tri-diagonal matrix $(M-2)^2$ times. They are defined as:

$A_1 = I - 2\tau \cdot$

$$\begin{bmatrix} -D_{2,2}-G_{2,2} & D_{2,2}+G_{2,2} & 0 & \cdot & \cdot \\ C_{3,2}-G_{3,2} & -C_{3,2}-D_{3,2} & D_{3,2}+G_{3,2} & 0 & \cdot \\ 0 & \cdot & \cdot & \cdot & \cdot \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ \cdot & \cdot & \cdot & \cdot & 0 \\ \cdot & \cdot & \cdot & \cdot & D_{M-2,N-1}+G_{M-2,N-1} \\ \cdot & \cdot & 0 & C_{M-1,N-1}-G_{M-1,N-1} & -C_{M-1,N-1}+G_{M-1,N-1} \end{bmatrix}$$

$A_2 = I - 2\tau \cdot$

$$\begin{bmatrix} -F_{2,2}-H_{2,2} & F_{2,2}+H_{2,2} & 0 & \cdot & \cdot \\ E_{2,3}-H_{2,3} & -E_{2,3}-F_{2,3} & F_{2,3}+H_{2,3} & 0 & \cdot \\ 0 & \cdot & \cdot & \cdot & \cdot \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ \cdot & \cdot & \cdot & \cdot & 0 \\ \cdot & \cdot & \cdot & \cdot & F_{M-1,N-2}+H_{M-1,N-2} \\ \cdot & \cdot & 0 & E_{M-1,N-1}-H_{M-1,N-1} & -E_{M-1,N-1}+H_{M-1,N-1} \end{bmatrix}$$

$$\underline{\bar{X}} = [X_{2,2}\ X_{3,2} \cdots X_{M-1,2}\ X_{2,3} \cdots X_{M-2,N-1}\ X_{M-1,N-1}]^T$$
$$\underline{\bar{\bar{X}}} = [X_{2,2}\ X_{2,3} \cdots X_{2,N-1}\ X_{3,2} \cdots X_{M-1,N-2}\ X_{M-1,N-1}]^T$$

Since both A_1 and A_2 are tri-diagonal matrices, one can get their inverses efficiently by using Thomas Algorithm([92]).

2.4 Experimental Results

In this section we present our experimental results on the application of the proposed model (2–9-2–10) to simulated data and a set of HARD MRI data from the human brain.

2.4.1 Analysis of simulated data

The aim of our experiment on the simulated data is to test whether our model can efficiently reconstruct a regularized ADC profile from the noisy HARD measurements. We simulated an ADC profile on a 2D lattice of size 8×4. The volume consists of two homogeneous regions, values of S_0 and all the $A_{l,m}$'s were shown in table 2-1.

Table 2-1. List of S_0 and $A_{l,m}$'s for two regions

Region	1	2
S_0	414	547
$A_{0,0}$	5.21×10^{-3}	1.43×10^{-2}
$A_{2,0}$	-1.17×10^{-3}	-2.68×10^{-3}
$ReA_{2,1}$	-4.37×10^{-5}	0
$ReA_{2,2}$	1.43×10^{-3}	0
$ImA_{2,1}$	3.64×10^{-5}	0
$ImA_{2,2}$	3.28×10^{-5}	0
$A_{4,0}$	-3.15×10^{-5}	8.4×10^{-6}
$ReA_{4,1}$	-1.56×10^{-4}	0
$ReA_{4,2}$	1.02×10^{-4}	0
$ReA_{4,3}$	6.30×10^{-5}	0
$ReA_{4,4}$	-8.54×10^{-5}	-1.73×10^{-3}
$ImA_{4,1}$	-8.01×10^{-5}	0
$ImA_{4,2}$	$0.9961.55 \times 10^{-4}$	0
$ImA_{4,3}$	1.41×10^{-5}	0
$ImA_{4,4}$	3.63×10^{-5}	0

In Figure 2.4.2 we displayed the true, noisy, and recovered ADC profiles $d(x, \theta, \phi)$ for the synthetic data with size 8×4. The ADC profile $d(x, \theta, \phi)$ was computed by (2–3) based on these simulated data, and the corresponding $s_{true}(x, \theta, \phi)$ was constructed via (1–6) with $b = 1000 s/mm^2$. Then the noisy HARD MRI signal $s(x, \theta, \phi)$ was generated by adding a zero mean Gaussian noise with standard deviation $\sigma = 0.15$. Figure 2.4.2b shows the ADC profile d computed by (2–3), where the coefficients of the SHS are the least-squares solutions of (1–11) with the noisy s.

We then applied our model (2–9-2–10) to the noisy $s(x, \theta, \phi)$ to test the effectiveness of the model, with $\lambda_{0,0} = 4, \lambda_{2,m} = 40(m = -2 \ldots 2), \lambda_{4,m} = 60(m = -4 \ldots 4)$. By solving the system of equations (2–14) in 2.5 seconds on computer with PIV 2.8GHZ CPU and 2G RAM using Matlab script code, we obtained 15 reconstructed functions as in (2–2). Using these $A_{l,m}$ (the solutions of (2–14)) we computed $d(x, \theta, \phi)$ via (2–3). The reconstructed $d(x, \theta, \phi)$ is shown in Figure 2.4.2c. Comparing these three figures, it is clear that the noisy measurements s have changed Figure. 2.4.2a, the original shapes of d, into Figure 2.4.2b. After applying our model (2–9-2–10) to reconstruct the ADC profiles, the shapes

of d in Figure 2.4.2a were recovered, as shown in Figure 2.4.2c. These simulated results demonstrate that our model is effective in simultaneously regularizing and recovering ADC profiles.

2.4.2 Analysis of human MRI data

The second test is to reconstruct and characterize ADC profiles $d(x, \theta, \phi)$ from human HARD MRI data.

The raw DWI data, usually contains a certain level of noise, were obtained on a GE 3.0 Tesla scanner using a single shot spin-echo EPI sequence. The scanning parameters for the DWI acquisition are: repetition time (TR)=1000ms, echo time (TE) =85ms, the field of view (FOV)=220 mm x 220 mm. 24 axial sections covering the entire brain with the slice thickness=3.8 mm and the intersection gap=1.2 mm. The diffusion-sensitizing gradient encoding is applied in fifty-five directions (selected for the HARD MRI acquisition) with $b = 1000 s/mm^2$. Thus, a total of fifty-six diffusion-weighted images, with a matrix size of 256 x 256, were obtained for each slice section. We applied model (2–9) to these data to compute the ADC profiles in the entire brain volume. By solving a system of equations (2–14) we obtained all the coefficients $A_{l,m}$'s in (2–2), and determined $d(x, \theta, \phi)$ using (2–3).

Then, we used these $A_{l,m}(x)$ to calculate R_0 and R_2 defined in (2–17) and (2–18) respectively, as well as the variance $\sigma(x)$ of $d(x, \theta, \phi)$ about its mean:

$$\sigma(x) = \int_0^{\pi} \int_0^{2\pi} (d(x, \theta, \phi) - \sum_{i=1}^{55} d(\mathbf{x}, \theta_i, \phi_i)/55)^2 d\theta d\phi.$$

Based on results from the HARD MRI data of this particular patient, we characterized the diffusion anisotropy according to the following procedure. If $R_0(x) > 0.856$, or $\sigma(\mathbf{x}) < 19.65$ the diffusion at x is classified as isotropic. For the remaining voxels if $R_2(x) > 0.75$, the diffusion at such voxels is considered as one-fiber diffusion. For uncharaterized voxels from these two steps we further classified them by the principles stated in the section 2.2. The selection of the thresholds mentioned above for R_0, R_2 and

σ involves experts' input and large sample experiments. Experimental results definitely depend on these thresholds, but not sensitively.

Figure. 2.6 presents $A_{2,0}(x)$, one of the coefficients in (2–3), for the particular slice in the volume. The images $A_{2,0}(x)$ in Figure. 2.6a and 2.6b are estimated by using (1–10) and solving (2–9), respectively.

Figure. 2-3 Compares FA and three $R_2(x)$'s with $A_{l,m}(\mathbf{x})$'s obtained from three different models for the same slice as shown in Figure.2.6 . Figure. 2-3a displays the FA image obtained by using advanced system software from GE. The $A_{l,m}(\mathbf{x})$'s used to obtain $R_2(\mathbf{x})$ in Figure. 2-3b are directly computed from (1–10). Those used to obtain $R_2(\mathbf{x})$ in Figures. 2-3c and 2-3d are the least-squares solutions of (1–11) and the solutions of (2–9), respectively. In Figures. 2-3c and 2-3d the voxels with high levels of intensities (red, yellow, yellow-light blue) are characterized as one-fiber diffusion.

Although the FA image in Figure. 2-3a is obtained based on a conventional DTI model (1–2), it is still comparable with the R_2 map, since single tensor diffusion characterized by SHS representation from the HARD images agrees with that characterized by the DTI model. However, in DTI a voxel with a low intensity of FA indicates isotropic diffusion, while using our algorithm, multi-fibers diffusion may occur at the location with the low value of R_2.

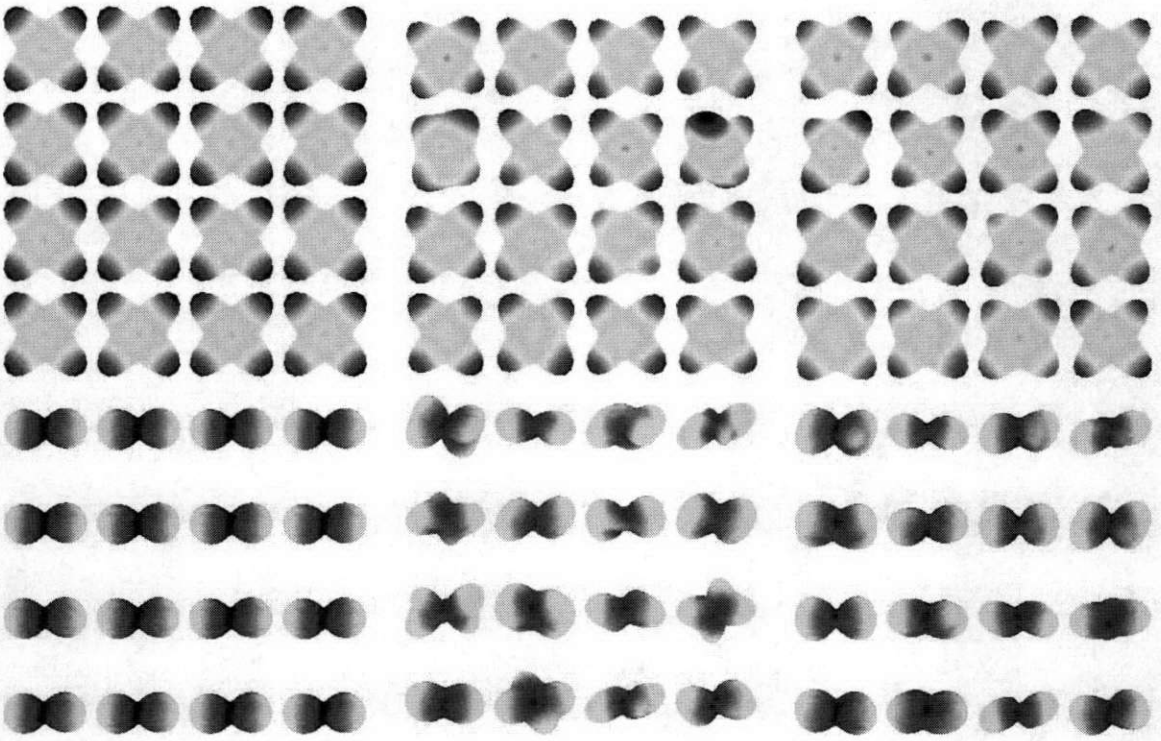
It is clearly evident that the ability to characterize anisotropic diffusion is enhanced, as shown in Figures. 2-3a-2-3d. Figure. 3b indicates again that the estimations of $A_{l,m}$ directly from the *log* signals usually are not good. Even the least-squares solution of (1–11) are not always effective. This can be seen by comparing the anatomic region inside the red square of Figures. 2-3c and 2-3d, which are zoomed in Figures. 2-4a and 2-4b, respectively. There is a dark broken line showing on the map of the external capsule (arrow to the right on Figure. 2-4a), this same region was recovered by the proposed model and characterized by the third step in our algorithm as two-fibers anisotropic diffusion (arrow to the right in Figure. 2-4b). (The model solutions reduced the value of

R_0, increased the values of R_1 slightly, and made the 3rd step in our characterization to be applied). Our results also showed the connection in a cortical associative tract (arrow to the left in Figures. 2-4b), however, this connection was not mapped out on Figure. 2-3c or the zoomed image in Figure. 2-4a. In fact this connection was not mapped out on Figures 2-3a-3b either. All these mapped connections are consistent with the known neuroanatomy. Combined together, our results indicate that our proposed model for joint recovery and smoothing of the ADC profiles has the advantage over the existing models for the enhancement of the ability to characterize the diffusion anisotropy.

Figure. 2-5a shows a partition of isotropic, one-fiber, and two-fiber diffusion for the same slice used in Figure 2-4. The two-fiber, one-fiber, and isotropic diffusion regions were further characterized by the white, gray, and black regions, respectively. The region inside the white square in Figure 2-5a, which is the same one squared in Figures. 2-3c and 2-3d, is zoomed in Figure. 2-4c. It is clearly to see the two arrayed voxels in Figure. 2-4b are classified as two-fiber diffusion. The characterization of the anisotropy on the voxels and their neighborhoods is consistent with the known fiber anatomy.

Figure. 2-5b represents the shapes of $d(x, \theta, \phi)$ at three particular voxels (upper, middle and lower rows). The d in all three voxels is computed using (2–3). However, the $A_{l,m}(\mathbf{x})$ used in computing d on the left column are the least-squares solutions of (1–11), while on the right column they are the solutions of the proposed model (2–9). The first and second rows show two voxels that can be characterized as isotropic diffusion before denoising, but as two-fibers diffusion after applying model (2–9). These two voxels are the same voxels as in Figure 2-4 directed by arrows. The lower row of Figure 2-5b shows the one-fiber diffusion was enhanced after applying our model.

Solving $A_{l,m}$'s of size $15 \times 109 \times 86 \times 8$ from 4-D data of size $55 \times 109 \times 86 \times 8$ takes 46.2 seconds for each iteration on computer with PIV 2.8GHZ CPU and 2G RAM in Matlab script code.



(a) (b) (c)

Figure 2-1. (a). True d, (b). The d generated by (2–3), where the $A_{l,m}$'s are the least square solution of (1–10) with the noisy measurement s, (c). Recovered d by applying model (2–9).

2.5 Existence

In this section we will discuss the existence of a solution to our minimization problem (2–9) using the idea developed in [86].

Recall that for a function $u \in BV(\Omega)$,

$$Du = \nabla u \cdot \mathcal{L}^n + D^s u$$

is a Radon measure, where ∇u is the density of the absolutely continuous part of Du with respect to the n-dimensional Lebesgue measure, $\mathcal{L}^n$, and $D^s u$ is the singular part. To minimize (2–9) over the functions in $BV(\Omega)$, we first need to give a precise definition for (2–9).

Definition: For $A_{l,m} \in BV(\Omega)$, define

$$\int_\Omega \phi_{l,m}(x, DA_{l,m}) := \int_\Omega \phi_{l,m}(x, \nabla A_{l,m})dx + \int_\Omega |D^s A_{l,m}|$$

where $\phi_{l,m}$ is defined as in (2–5), and $\int_\Omega |D^s A_{l,m}|$ is the total variation norm of $A_{l,m}$.

Then, our energy functional (2–9) is defined as

$$E(A_{l,m}) = \lambda \int_\Omega \sum_{l=0,2,4} \sum_{m=-l}^{l} \phi_{l,m}(x, \nabla A_{l,m}) + \lambda \int_\Omega |D^s A_{l,m}|$$

$$+ \frac{1}{2} \int_\Omega \int_0^{2\pi} \int_0^{\pi} |s(x,\theta,\phi) - s_0(x) e^{-bd(x,\theta,\phi)}|^2 sin\theta d\theta d\phi dx. \qquad (2\text{–}23)$$

In the discussion of existence, without loss of generality, we set the parameter $\lambda = 1$ in (2–9) and threshold $M_{l,m} = 1$ in (2–5) to reduce the complexity in the formulation.

Next we will show the lower semi-continuity of the energy functional (2–9) in L^1, i.e. if for each l, m ($l = 0, 2, 4$ and $m = -l, \ldots, l$), as $k \to \infty$,

$$A^k_{l,m} \to A^0_{l,m} \quad in \quad L^1(\Omega),$$

then

$$E(A^0_{l,m}) \leq liminf_{k\to\infty} E(A^k_{l,m}), \qquad (2\text{–}24)$$

To prove this we need the following lemma:

Lemma: Let

$$\phi(x,r) := \begin{cases} \frac{1}{p(x)}|r|^{p(x)}, & |r| \le 1 \\ |r| - (1 - \frac{1}{p(x)}), & |r| > 1 \end{cases} \tag{2–25}$$

For $u \in BV(\Omega)$ denote

$$\Phi(u) := \int_\Omega \phi(x, Du),$$

and

$$\widetilde{\Phi}(u) := \sup_{\substack{\psi \in C_0^1(\Omega, R^n) \\ |\psi| \le 1}} \int_\Omega -u div\psi - \frac{p(x)-1}{p(x)} |\psi|^{\frac{p(x)}{p(x)-1}} dx.$$

Then,

$$\Phi(u) = \widetilde{\Phi}(u) \tag{2–26}$$

Furthermore, $\Phi(u)$ is lower semi-continuous on $L^1(\Omega)$, i.e. if $u_j, u \in BV(\Omega)$ satisfy $u_j \to u$ weakly in $L^1(\Omega)$ as $j \to \infty$ then

$$\Phi(u) \le \liminf_{j \to \infty} \Phi(u_j).$$

Proof First note that for each $\psi \in C_0^1(\Omega, R^n)$, the map

$$u \longrightarrow \int_\Omega -u div\psi - \frac{p(x)-1}{p(x)} |\psi|^{\frac{p(x)}{p(x)-1}} dx$$

is continuous and affine on $L^1(\Omega)$. Therefore, $\widetilde{\Phi}(u)$ is convex and lower semi-continuous on $L^1(\Omega)$ and the domain of $\widetilde{\Phi}(u)$, $\{u \,|\, \widetilde{\Phi}(u) < \infty\}$, is precisely $BV(\Omega)$.

Next we show (2–26). For $u \in BV(\Omega)$, we have that for each $\psi \in C_0^1(\Omega, R^n)$,

$$-\int_\Omega u div\psi dx = \int_\Omega \nabla u \cdot \psi dx + \int_\Omega D^s u \cdot \psi$$

and so

$$\widetilde{\Phi}(u) = \sup_{\substack{\psi \in C_0^1(\Omega, R^n) \\ |\psi| \le 1}} \int_\Omega \nabla u \cdot \psi - \frac{p(x)-1}{p(x)} |\psi|^{\frac{p(x)}{p(x)-1}} dx + \int_\Omega D^s u \cdot \psi$$

3005

Since the measures dx and $D^s u$ are mutually singular, by a standard argument we can have

$$\widetilde{\Phi}(u) = \sup_{\substack{\psi \in C_0^1(\Omega, R^n) \\ |\psi| \leq 1}} \int_\Omega (\nabla u \cdot \psi - \frac{p(x)-1}{p(x)} |\psi|^{\frac{p(x)}{p(x)-1}}) dx + \int_\Omega |D^s u|.$$

To prove (2–26) it only remains to show that

$$\int_\Omega \phi(x, \nabla u) dx = \sup_{\substack{\psi \in C_0^1(\Omega, R^n) \\ |\psi| \leq 1}} \int_\Omega (\nabla u \cdot \psi - \frac{p(x)-1}{p(x)} |\psi|^{\frac{p(x)}{p(x)-1}}) dx. \tag{2–27}$$

Since any $\rho \in L^\infty(\Omega, R^n)$ can be approximated in measure by $\psi \in C_0^1(\Omega, R^n)$, we have that

$$\sup_{\substack{\psi \in C_0^1(\Omega, \mathbb{R}^n) \\ |\psi| \leq 1}} \int_\Omega \nabla u \cdot \psi - \frac{p(x)-1}{p(x)} |\psi|^{\frac{p(x)}{p(x)-1}} dx = \sup_{\substack{\rho \in L^\infty(\Omega, R^n) \\ |\rho| \leq 1}} \int_\Omega \nabla u \cdot \rho - \frac{p(x)-1}{p(x)} |\rho|^{\frac{p(x)}{p(x)-1}} dx. \tag{2–28}$$

Choosing $\rho(x) = 1_{\{|\nabla u| \leq 1\}} |\nabla u|^{p(x)-1} \frac{\nabla u}{|\nabla u|} + 1_{\{|\nabla u| > 1\}} \frac{\nabla u}{|\nabla u|}$, where 1_E is the indicator function on E, we see that the right hand side of (2–28) is

$$\geq \int_\Omega \frac{1}{p(x)} |\nabla u|^{p(x)} 1_{\{|\nabla u| \leq 1\}} + \left[|\nabla u| - \frac{p(x)-1}{p(x)} \right] 1_{\{|\nabla u| > 1\}} dx = \int_\Omega \phi(x, \nabla u) dx. \tag{2–29}$$

To show the opposite inequality, we argue as follows. For any $\rho \in L^\infty(\Omega, R^n)$, since $p(x) > 1$ we have that for almost all x,

$$\nabla u(x) \cdot \rho(x) \leq \frac{1}{p(x)} |\nabla u|^{p(x)} + \frac{p(x)-1}{p(x)} |\rho(x)|^{\frac{p(x)}{p(x)-1}}$$

In particular, if $|\nabla u| \leq 1$,

$$\nabla u(x) \cdot \rho(x) - \frac{p(x)-1}{p(x)} |\rho(x)|^{\frac{p(x)}{p(x)-1}} \leq \frac{1}{p(x)} |\nabla u|^{p(x)} \tag{2–30}$$

If $|\nabla u| > 1$, noticing $p(x) > 1$ and $|\rho| \leq 1$ for almost all x we have that

$$\nabla u \cdot \rho = |\nabla u| \frac{\nabla u}{|\nabla u|} \cdot \rho \leq |\nabla u| \left[\frac{1}{p(x)} + \frac{p(x)-1}{p(x)} |\rho|^{\frac{p(x)}{p(x)-1}} \right]$$

and so

$$\nabla u\cdot\rho-\frac{p(x)-1}{p(x)}|\rho|^{\frac{p(x)}{p(x)-1}}\leq\frac{1}{p(x)}|\nabla u|+(|\nabla u|-1)\frac{p(x)-1}{p(x)}|\rho|^{\frac{p(x)}{p(x)-1}}\leq|\nabla u|-\frac{p(x)-1}{p(x)} \quad (2\text{–}31)$$

Combining, (2–28), (2–29), (2–30), and (2–31), we have (2–27), and hence for all $u\in BV(\Omega)$, $\widetilde{\Phi}(u)=\Phi(u)$.

Note that $\phi(x,r)=\phi_{l,m}(x,r)$, if $p(x)=p_{l,m}(x)$. As a direct consequence of this lemma we have E_1 in (2–4) is weakly lower semi-continuous on $L^1(\Omega)$.

Furthermore, we can show that E_2 in (2–4) is lower semi-continuous on $L^1(\Omega)$. Indeed, when

$$A^k_{l,m}\rightarrow A^0_{l,m},\ \ in\ \ L^1(\Omega),\ \ as\ \ k\rightarrow\infty,$$

for all $l=0,2,4$ and $m=-l,\ldots,l)$, $A^k_{l,m}\rightarrow A^0_{l,m}$ *a.e.* on Ω. Then, if $s(x,\theta,\phi)\in L^2(\Omega\times S^2)$ and $s_0(x)\in L^2(\Omega)$, by the dominated convergence theorem, we have

$$E_2(A^k_{l,m})\rightarrow E_2(A^0_{l,m}).$$

Therefore, $E=E_1+E_2$ is lower semi-continuous on $L^1(\Omega)$, and (2–24) holds.

Now we can prove the existence results.

Theorem: Let Ω be a bounded open set of R^n. Assume that $s(x,\theta,\phi)\in L^2(\Omega\times S^2)$ and $s_0(x)\in L^2(\Omega)$. Then, there exists a solution consisting of functions $A^0_{l,m}$ ($l=0,2,4$ and $m=-l,\ldots,l$) to the minimization problem (2–9) over the space of $BV(\Omega)$.

Proof: Let $A^k_{l,m}$ ($l=0,2,4$ and $m=-l,\ldots,l$) be the minimizing sequences of (2–9) in $BV(\Omega)$. Then for each (l,m) the sequence $A^k_{l,m}$ is bounded in $BV(\Omega)$. From the compactness of $BV(\Omega)$ there exist subsequences of $A^k_{l,m}$ (still denoted by $A^k_{l,m}$) and functions $A^0_{l,m}\in BV(\Omega)$ satisfying

$$A^k_{l,m}\rightarrow A^0_{l,m}\ \ strongly\ \ in\ \ L^1(\Omega).$$

By the lower semi-continuity of the energy functional on $L^1(\Omega)$, (see (2–24)), we have

$$E(A^0_{l,m}) \leq liminf_{k\to\infty} E(A^k_{l,m}) \leq inf_{A_{l,m}\in BV(\Omega)} E(A_{l,m}).$$

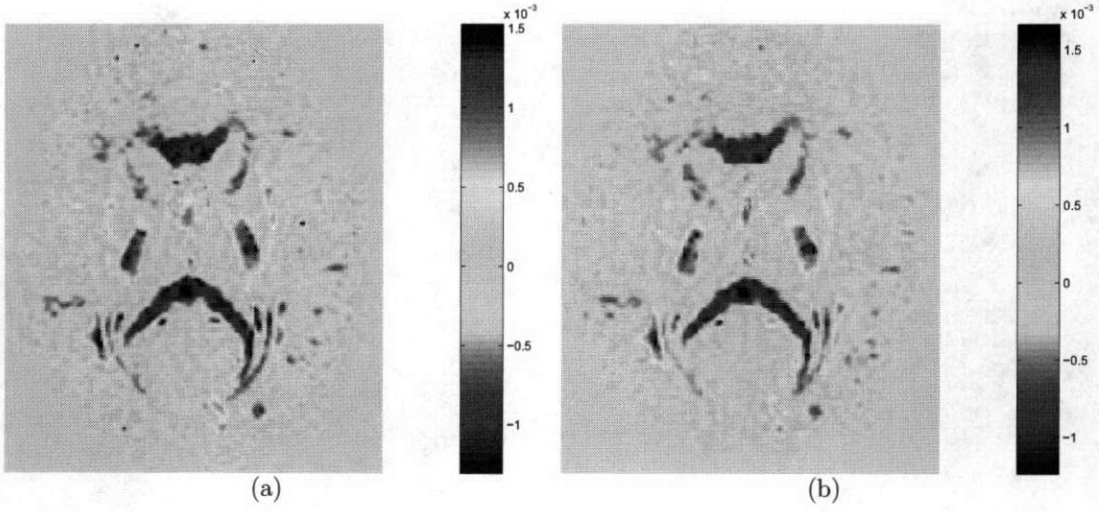
Hence, all these $A^0_{l,m}$ ($l = 0, 2, 4$ and $m = -l, \ldots, l$) together form a solution to the minimization problem (2–9).

2.6 Conclusions

A novel variational framework was introduced for simultaneous smoothing and estimation of ADC profiles in the form of truncated SHS based on HARD MRI. This model is featured by minimizing a nonstandard growth function with nonlinear data fitting. Moreover, the constraints on the positiveness and antipodal symmetry properties of d is also accommodated in the model. We also demonstrated our algorithm for using the variance of d from its mean and the coefficients of its truncated SHS approximation to characterize the diffusion anisotropy.

Our experiments on both synthetic data and human HARD MRI data showed the effectiveness of the proposed model in the estimation of ADC profiles and the enhancement of the characterization of diffusion anisotropy. The characterization of non-Gaussian diffusion from the proposed method is consistent with the known neuroanatomy.

The choice of the current parameters, however, may affect the results. Our choice is made based on the principle that classification for one-fiber diffusion from the model solution should agree with a priori knowledge of the fiber connections. In this article, we have not included the work for determination of fiber directions and the method for automated fiber tracking. The study addressing these problems will be reported in separate papers.



(a) (b)

Figure 2-2. (a). A_{20} computed from (1–10), (b). A_{20} obtained from model(2–9)

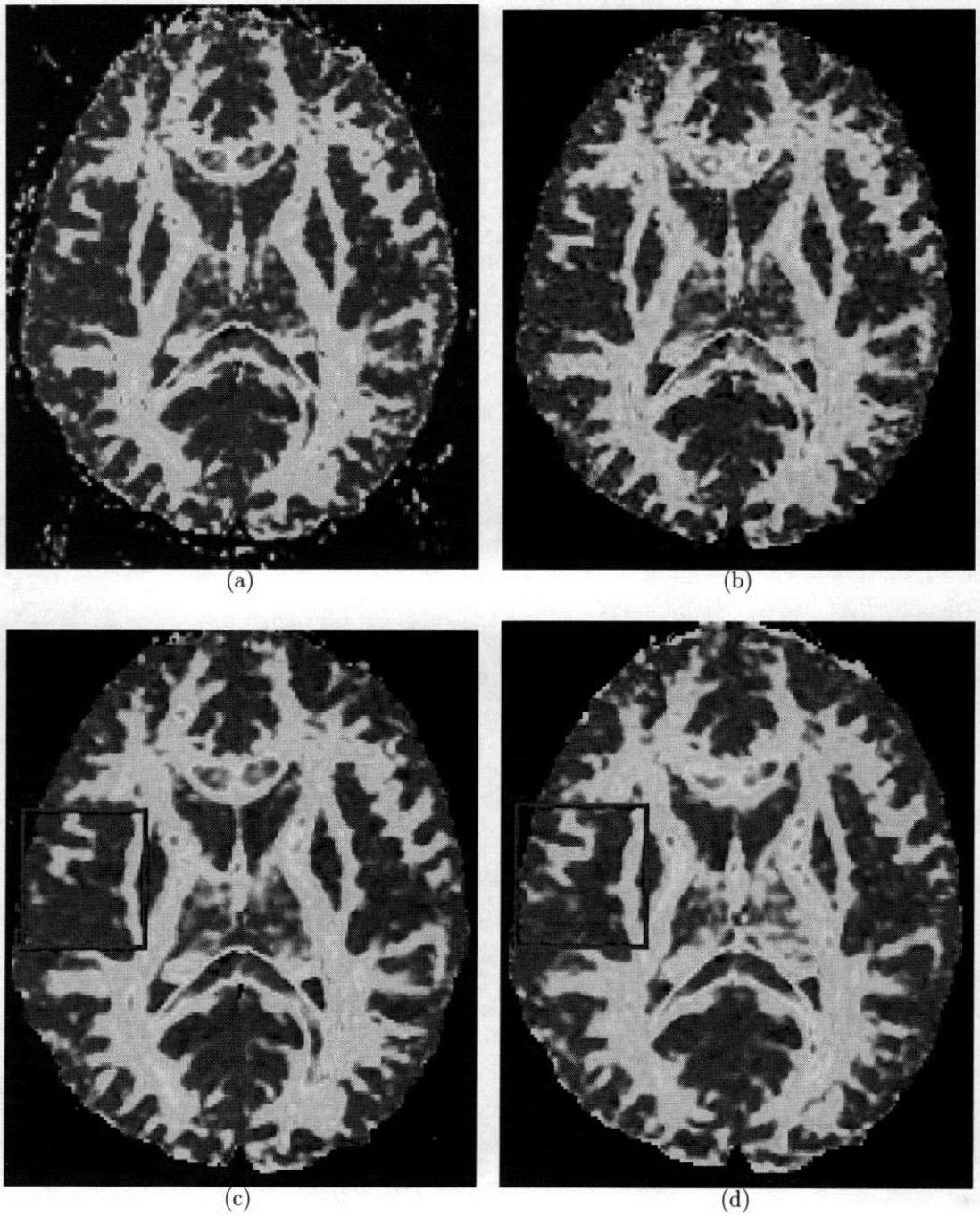
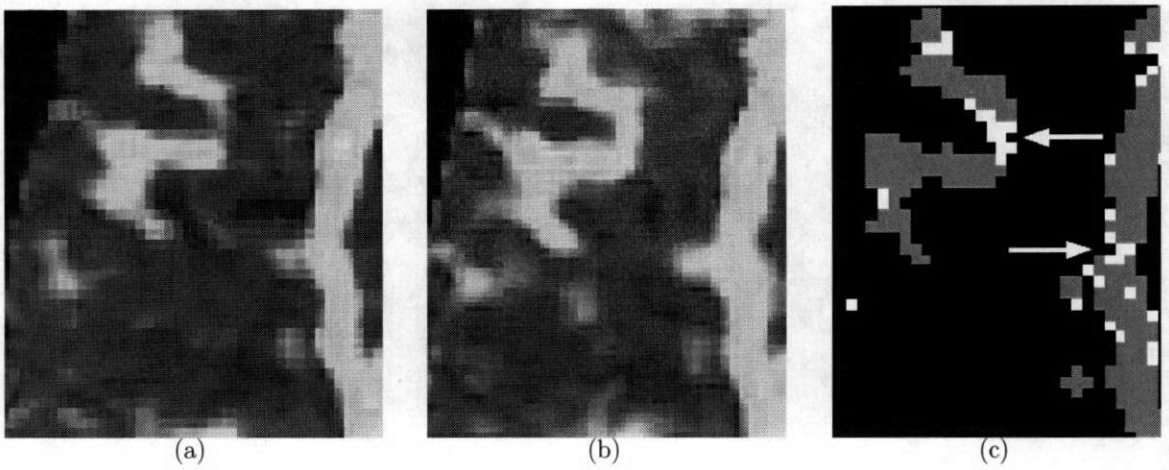
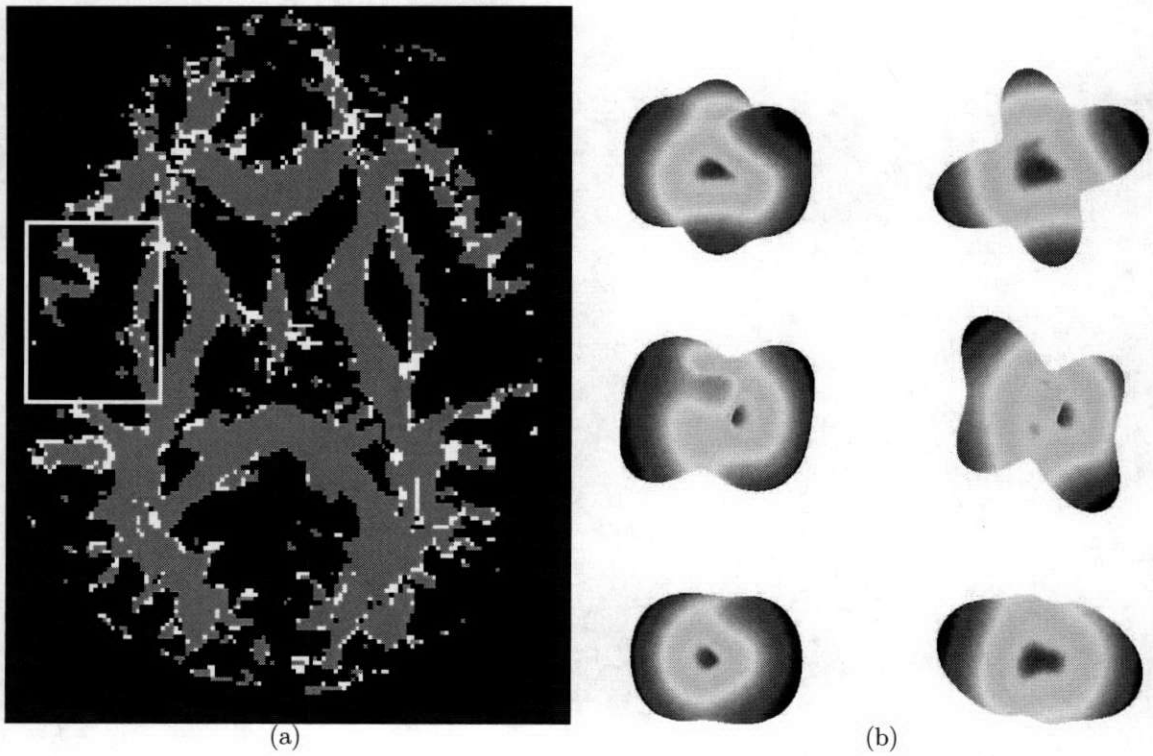


Figure 2-3. (a). FA from GE software, (b)-(d). R_2 with the $A_{l,m}$'s as the solutions of (1–10), least-squares solutions of (1–11), and model solutions, respectively.



(a) (b) (c)

Figure 2-4. (a)-(b). Enlarged portions inside the red squares in Figures 2-3c and 2-3d, respectively. (c). Enlarged portions inside the white squares in Figure 2-5a.



(a) (b)

Figure 2-5. (a).Classification: white, gray, and black voxels are identified as two-fiber, one-fiber, and isotropic diffusion respectively, (b). Shapes of $d(x, \theta, \phi)$ at three particular points (upper, middle and lower rows). The d is computed via (2–3). $A_{l,m}(\mathbf{x})$ used in (2–3) on the left columns are the least-squares solutions of (1–11), while on the right column are the model solutions.

CHAPTER 3
ESTIMATION,SMOOTHING AND CHARACTERIZATION OF ADC IN HARD WEIGHTED IMAGE USING PRODUCT OF TWO SPHERICAL HARMONIC SERIES UP TO ORDER 2

3.1 New Approximation Model For ADC Profiles

In[14, 21, 93], to detect the diffusion with at most two fiber orientation, the ADC profiles were represented by a truncated SHS up to order 4 in the form of (1–9). In [14] the coefficients $A_{l,m}$'s (l is even) were determined by inverse spherical harmonic transform of $-\frac{1}{b}log\frac{s(\mathbf{q})}{s_0}$ and in [21] they were estimated as the least-squares solutions of

$$-\frac{1}{b}log\frac{s(\mathbf{q})}{s_0} = \sum_{l=0}^{l_{max}} \sum_{m=-l}^{l} A_{l,m} Y_{l,m}(\theta,\phi). \tag{3–1}$$

Regularization on the raw data or $A_{l,m}$ wasn't considered in these two work. In [93] $A_{l,m}$'s were considered as a function of $\mathbf{x}$, and estimated and smoothed simultaneously by solving a constrained minimization problem:

$$\min_{A_{l,m}(\mathbf{x}),\tilde{s}_0(\mathbf{x})} \int_\Omega \{ \sum_{l=0,2,4} \sum_{m=-l}^{l} |\nabla A_{l,m}(\mathbf{x})|^{p_{l,m}(\mathbf{x})} + |\nabla \tilde{s}_0(\mathbf{x})|^{p(\mathbf{x})} \} d\mathbf{x}$$

$$+ \frac{\lambda}{2} \int_\Omega \{ \int_0^{2\pi} \int_0^{\pi} |s(\mathbf{x},\mathbf{q}) - \tilde{s}_0(\mathbf{x}) e^{-bd(\mathbf{x},\theta,\phi)}|^2 sin\theta d\theta d\phi + |\tilde{s}_0 - s_0|^2 \} d\mathbf{x}, \tag{3–2}$$

with the constraint $d > 0$. In this model $p_{l,m}(\mathbf{x}) = 1 + \frac{1}{1+k|\nabla G_\sigma * A_{l,m}|^2}$, $q(\mathbf{x}) = 1 + \frac{1}{1+k|\nabla G_\sigma * s_0|^2}$, and d takes the form (1–9). By the choice of $p_{l,m}$ and q, the regularization is total variation based near edges, isotropic in homogeneous regions, and between isotropic and total variation based that varies depending on the local properties of the image at other locations, In these work since the ADC profile was approximated by (1–9), at least 15 measurements of $s(q)$ were required to to estimate the 15 coefficients $A_{l,m}$.

However, the mixture model (1–4) with $n = 2$, which is also able to detect two-fiber diffusion involves only 13 unknown functions. This motivates us to find a model that is

able to detect non-Gaussian diffusion with the minimum number of unknowns. In this chapter, we only discuss the diffusion with no more than two fiber orientations within a voxel. The significance of this study is clear: less number of unknowns lead to less requirement for number of HARD measurements. This will significantly reduce the scan time and thus is important in clinical application.

Our basic idea is to approximate the ADC profiles MRI by the product of two second order SHS's instead of a SHS up to order four. This can be formulated as

$$d(\mathbf{x},\theta,\phi) = (\sum_{l=0,2}\sum_{m=-l}^{l} b_{l,m}(\mathbf{x})Y_{l,m}(\theta,\phi)) \cdot (\sum_{l=0,2}\sum_{m=-l}^{l} c_{l,m}(\mathbf{x})Y_{l,m}(\theta,\phi)). \tag{3–3}$$

In this model there are only 12 unknowns: $b_{l,m}$, $c_{l,m}$ ($l = 0, 2$ and $-l \leq m \leq l$).

To estimate the ADC profile from the raw HARD MRI data, which usually contains a certain level of noise, we propose a simultaneous smoothing and estimation model similar to (3–2) for solving $b_{l,m}$, $c_{l,m}$, that is the following constrained minimization problem:

$$\min_{b_{l,m}(\mathbf{x}),c_{l,m}(\mathbf{x}),\tilde{s}_0(\mathbf{x})} \int_\Omega \{\sum_{l=0,2}\sum_{m=-l}^{l} \alpha(|\nabla b_{l,m}(\mathbf{x})| + |\nabla c_{l,m}(\mathbf{x})| + \beta|\nabla \tilde{s}_0(\mathbf{x})|d\mathbf{x}$$

$$+ \frac{1}{2}\int_\Omega \{\int_0^{2\pi}\int_0^{\pi} |s(\mathbf{x},\mathbf{q}) - \tilde{s}_0(\mathbf{x})e^{-bd(\mathbf{x},\theta,\phi)}|^2 sin\theta d\theta d\phi + |\tilde{s}_0 - s_0|^2\}d\mathbf{x}, \tag{3–4}$$

with constraint $d \geq 0$, where d is in the form of (3–3). α, β are constants. The first 3 terms are the regularization terms for $b_{l,m}$, $c_{l,m}$ and s_0 respectively. The last two terms are the data fidelity terms based on the original Stejskal-Tanner equation(1–6).

Next, feasibility of this model will be explained. Denote $B := \sum_{l=0,2}\sum_{m=-l}^{l} b_{l,m} \cdot Y_{l,m}(\theta,\phi)$, $C := \sum_{l=0,2}\sum_{m=-l}^{l} c_{l,m}Y_{l,m}(\theta,\phi)$ and $A := \sum_{l=0,2,4}\sum_{m=-l}^{l} A_{l,m}Y_{l,m}$. Define sets $SBC = \{d : d(\theta,\phi) = B \cdot C\}, SA = \{d : d(\theta,\phi) = A\}$. Since each $d(\theta,\phi)$ in SBC is a function defined on S^2, it can be approximated by SHS, simple calculation shows that coefficients of the approximated SHS of even order larger than 4 are all zeros, so $SBC \subset SA$. On the other hand,numerous experiments show that when a voxel is not more complicated than 2-fiber diffusion, its ADC is always a function in set SBC. But if a voxel

is of 3-fiber diffusion or even more complicated, its ADC can not be described accurately by a function in SBC. This implies that SBC is a real subset of SA. Fig.3-1 depicts how functions in set SBC and SA differ in representing ADC. It is observed that the ADC in the form A can not be well approximated by $B \cdot C$ only in 3-fiber diffusion case (see Fig.3-1). Therefore, if we focus only on characterizing at most two-fiber diffusion, which is the most interesting case, model (3–3) is reasonable and sufficient to represent ADC.

Model (3–4) is a minimization problem with constraint $d(\theta,\phi) \geq 0$ for all $0 \leq \theta < \pi, 0 \leq \phi < 2\pi$ which is usually difficult to implement. To improve the efficiency of computation we used the idea that any second order SHS $\sum_{l=0,2}\sum_{m=-l}^{l} b_{l,m}Y_{l,m}(\theta,\phi)$ is equivalent to a tensor model u^TDu for some semi-positive definite 3×3 matrix D, where $u(\theta,\phi) = (sin\theta cos\phi, sin\theta sin\phi, cos\theta)$. This means that the coefficients $b_{l,m}$,$(l = 0,2,m = -l,...,l)$ in SHS and the entries $D(i,j)$, $(i,j = 1,...,3)$ in D can be computed from each other explicitly. Here are two examples: $b_{00} = \frac{2}{3}\sqrt{\pi}(D_{11} + D_{22} + D_{33})$. $D(1,1) = -\frac{\sqrt{5}b_{20} - 2b_{00} - \sqrt{30}Re(b_{22})}{4\sqrt{\pi}}$, where $Re(b_{22})$ is the real part of b_{22}. Hence, we could let $B = uD_1u^T$,$C = uD_2u^T$, then $d = (uD_1u^T)(uD_2u^T)$. And for $i = 1,2$ decomposed D_i into $D_i = L_iL_i^T$ with L_i a lower triangular matrix to guarantee semi-positiveness of D_i. The ADC is finally approximated by

$$d(\mathbf{x},\theta,\phi) = [u(\theta,\phi)L_1(\mathbf{x})L_1(\mathbf{x})^Tu(\theta,\phi)^T][(u(\theta,\phi)L_2(\mathbf{x})L_2(\mathbf{x})^Tu(\theta,\phi)^T]. \tag{3–5}$$

Furthermore we substituted model (3–4) by

$$\min_{L_1^{jk}(\mathbf{x}),L_2^{jk}(\mathbf{x}),\tilde{s}_0(\mathbf{x})} \int_\Omega (\alpha \sum_{i=1}^{2}\sum_{j=1}^{3}\sum_{k=1}^{j} |\nabla L_i^{j,k}| + \beta|\nabla \tilde{s}_0|)d\mathbf{x}$$

$$\frac{1}{2}\int_\Omega \{\int_0^{2\pi}\int_0^{\pi} |s - \tilde{s}_0e^{-bd}|^2 sin\theta d\theta d\phi + |\tilde{s}_0 - s_0|^2\}d\mathbf{x}, \tag{3–6}$$

where $d = (uL_1L_1^Tu^T)(uL_2L_2^Tu^T)$. All the $b_{l,m}, c_{l,m}; l = 0,2,m = -l...l$ are smooth functions of L_i^{jk},$i = 1,2; j = 1,2,3; k \leq j$, smoothness of L_i^{jk} guarantees that of $b_{l,m}$'s, $c_{l,m}$'s. The first term in model (3–6) thus works equivalently as the first two terms in

model (3–4) do, while all the other terms are the same as those left in (3–4). Hence, (3–6) is equivalent to (3–4), but it is a non-constrain minimization problem and is thus easy to implement. After we get L_1 and L_2, $b_{l,m}$ and $c_{l,m}$ in (3–4) can be obtained by the one to one relation between them.

(a) (b) (c) (d)

Figure 3-1. Comparison of the ADC's approximated by (1–9) and (3–3) in four cases: (a) isotropic diffusion, (b) one-fiber diffusion, (c) two-fiber diffusion, (d) three-fiber diffusion. In (a)-(d) from left to right, top to bottom, we show shapes of B, C, $B \cdot C$, and A, respectively.

We apply model(3–6) to a set of human brain HARD MRI data to reconstruct and characterize ADC profiles. The data set consists of 55 diffusion weighted images $S_k : \Omega \to R, k = 1, ..., 55$, and one image S_0 in the absence of a diffusion-sensitizing field gradient(b=0 in (1–6)). 24 evenly spaced axial planes with 256×256 voxels in each slice are obtained using a 3T MRI scanner with single shot spin-echo EPI sequence. Slice thickness is $3.8mm$, gap between two consecutive slices is $1.2mm$, repetition time $(TR) = 1000ms$, echo time $(TE) = 85ms$ and $b = 1000s/mm^2$. The field of view (FOV) $=220mm \times 220mm$. We first applied the model(3–6) to the data to get L_i, and then used L_i to compute $b_{l,m}$ and $c_{l,m}$, $l = 0, 2, m = -l..l$, and the ADC $d = B \cdot C$. On the other hand, we used the model (3–2) to estimate $A_{l,m}$ and get A. The comparison for the shapes of ADC in the form of $B \cdot C$ and A is demonstrated in Fig.3-3(a)-(d) at four specific voxels. The diffusion at these 4 voxels are isotropic (a), one-fiber (b), two-fiber (c), and three-fiber (d), respectively. In each sub figure, the up left, up right, down left, down right ones are the shapes of B, C, $B \cdot C$ and A, respectively. It is evident that if the diffusion is isotropic,

one-fiber or two-fiber, $B \cdot C$ and A are the same. However, if the diffusion is three-fiber, A can't be well approximated by $B \cdot C$.

To show the effectiveness of the proposed model in recovering ADC, in Fig.3-2(a)-(d) we compared images of R_2 (defined in section 3.2) with coefficient $A_{l,m}$ estimated by 4 different methods. The voxels with higher value of R_2 were considered as one-fiber diffusion. The $A_{l,m}$'s in (a), (b) and (c) were estimated using least-squares method in [21], model (3–2), and model(3–6) with the diffusion-sensitizing gradient applied to 55 directions, respectively. The $A_{l,m}$'s in (d) are estimated by the same way as that in (c), but from the HARD data with 12 carefully chosen directions. The model (3–6) applied on 55 measurements worked as good as the model (3–2) in getting higher value of R_2. Both of them worked better than the least-squares method that does not consider regularization. Although the result from 12 measurements was not as good as that from 55 measurements, they are are still comparable. We will show in Fig.3-5(a) and (b) that the anisotropy characterization results based on the ADC presented in (c) and (d) are also close. These experimental results indicated that by using the proposed model the voxels with two-fiber diffusion can be detected reasonably well from 12 HARD measurements in carefully selected directions.

3.2 Use Of CRE To Characterize Anisotropy

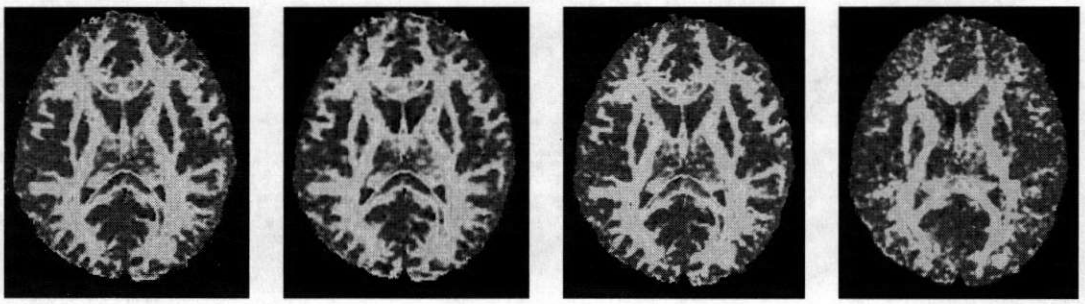
As mentioned, FA is only able to detect Gaussian diffusion. For non-Gaussian diffusion, Frank and Alexander et.al. used the order of significant component in SHS to characterize anisotropy. They considered voxels with significant 4th order components as two-fiber diffusion. In [93] Chen et al. realized that such a voxel could have isotropic or one-fiber diffusion. They defined $R_0 := \frac{|A_{0,0}|}{\sum_{l=0,2,4}\sum_{m=-l}^{l}|A_{l,m}|}$, $R_2 := \frac{\sum_{m=-2}^{m=2}|A_{2,m}|}{\sum_{l=0,2,4}\sum_{m=-l}^{l}|A_{l,m}|}$. Higher values of R_0 and R_2 are corresponding to isotropic and one-fiber diffusion, respectively. For the rest of points, the number of local maxima of ADC, together with the weights of the variances at the local maxima were used to classify voxels as isotropic, one-fiber or two-fiber diffusion. This procedure is more precise, but there are many measures involved

and thus more thresholds needed to be set subjectively. In this section, we will introduce a simple scheme using only one measurement CRE and two thresholds.

CRE is a measure of uncertainty/information in a random variable. Let X be a random variable in R, CRE of X is defined by

$$CRE(X) = -\int_{R_+} P(X > \lambda) log P(X > \lambda) d\lambda, \tag{3–7}$$

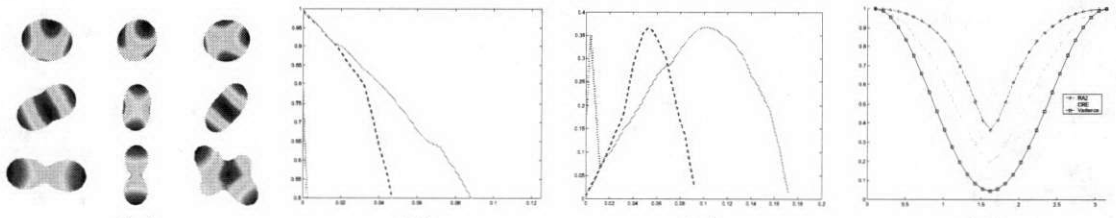
where $R_+ = \{X \in R | X \geq 0\}$.



(a) (b) (c) (d)

Figure 3-2. (a)-(d) are images of R_2 with $A_{l,m}$'s calculated using least-squares method, model (3–2), model (3–6) applied on 55 measurements, and model (3–6) applied on 12 measurements, respectively.

We use CRE of e^{-bd} rather then d to characterize diffusion anisotropy, where d is recovered from HARD measurements through(3–6). The magnitude of ADC is usually in



(a) (b) (c) (d)

Figure 3-3. (a) Shapes of ADC at an isotropic (first row),one-fiber (second row) and two-fiber (last row). (b)-(c) Graphs of $F(\lambda)$, $-F(\lambda) log F(\lambda)$ at three particular voxels:isotropic (red),one-fiber (green),two-fiber (blue).(d) R_2 (blue),CRE (yellow),variance (black) as functions of rotation angle ψ used in constructing synthetic data.

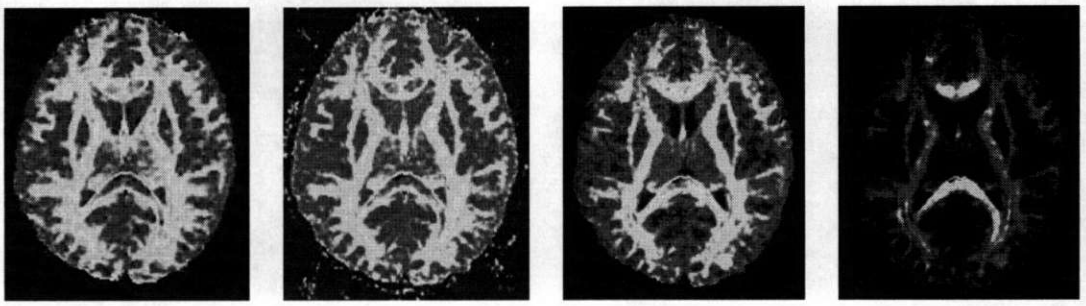
the order of 10^{-3}, while the magnitude of e^{-bd} is in the order of 10^{-1}, which is larger than that of ADC itself. Moreover, e^{-bd} is a smooth approximation of the data s/s_0.

The weak convergence property of CRE proved in [93] makes empirical CRE computation based on the samples converges in the limit to the true CRE. This is not the case for the Shannon entropy. We define empirical CRE of e^{-bd} as

$$CRE(e^{-bd}) = -\sum_{i=2}^{M} P(e^{-bd} > \lambda_i) log P(e^{-bd} > \lambda_i) \Delta\lambda_i \tag{3-8}$$

where $\{\lambda_1 < \lambda_2 < ... < \lambda_M\}$ is range of e^{-bd} at voxel $\mathbf{x}$. $\Delta\lambda_i = \lambda_i - \lambda_{i-1}$ is the absolute difference between two adjacent e^{-bd}, note this term is not shown in Shannon entropy. In most of the cases, the variation of e^{-bd} is the largest for one-fiber diffusion voxels, smaller for two-fiber diffusion and smallest for isotropic voxels. This also explains why CRE is the largest for one-fiber, medium for two-fiber and smallest for isotropic diffusion voxels. In our experiment, we choose M=1000 uniformly distributed directions (θ, ϕ) in (3-8).

Define the decreasing distribution function $F(\lambda) := P(e^{-bd} > \lambda)$. Fig.3-3(b) shows the graphs of $F(\lambda)$ at three pre-classified voxels: isotropic (red), one-fiber(green), two-fiber(blue). It is observed that the support and magnitude of $F(\lambda)$ are largest at the voxel with one-fiber diffusion, and smallest at that with isotropic diffusion. Fig.3-3 (c) demonstrate the graphs of $-F(\lambda)logF(\lambda)$ at the same three voxels. It is evident that the area under the green curve(one-fiber) is much larger than that under the blue curve(two-fiber), while the area under the red curve is the smallest. Since CRE is exactly the area under curve $-F(\lambda)logF(\lambda)$, we can conclude that measure $CRE(e^{-bd})$ is the largest at the voxels with one-fiber diffusion, medium with two-fiber diffusion, and smallest with isotropic diffusion. Thus measure $CRE(e^{-bd})$ could be used to discern isotopic, one-fiber and two-fiber diffusion with two thresholds T_1 and T_2, with $T_1 < T_2$. Set up 3 intervals:$(0, T_1)$, (T_1, T_2), (T_2, ∞). Voxels with CRE fall into the first, second, and third intervals are classified as isotropic, two-fiber and one-fiber diffusion respectively.



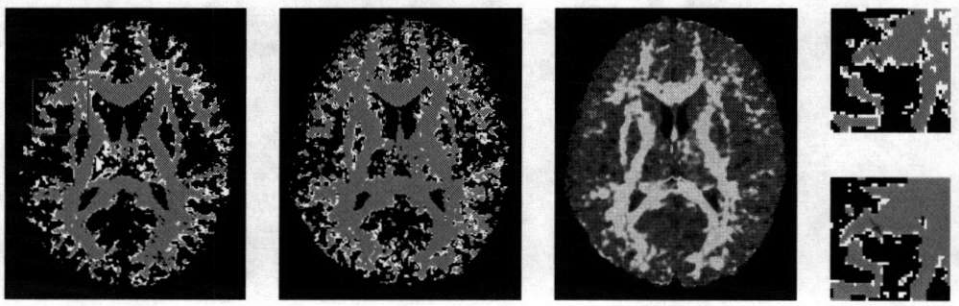
(a) (b) (c) (d)

Figure 3-4. Images of four measures: (a)R_2, (b) FA, (c) CRE of e^{-bd}, (d) Variance of e^{-bd}.

Fig.3-3(d) on synthetic data and Fig.3-4 on human brain HARD MRI data further show the strengths of CRE over the three popularly used measures R_2, FA and variance in characterizing diffusion anisotropy. The human data is the same as that used in Fig.3-2. The synthetic data is constructed as follows: Set D_1 and D_2 to be two diagonal matrix with diagonal element $4 \times 10^{-2}, 10^{-2}, 2 \times 10^{-2}$ and $8 \times 10^{-2}, 10^{-2}, 3 \times 10^{-2}$, respectively. Then fix D_1 but rotate principle eigenvector of D_2 by angle ψ to get $D_2(\psi)$ Let $B(\theta, \phi) = u^T D_1 u$, $C_\psi(\theta, \phi) := u^T D_2(\psi) u$. We computed R_2, FA and CRE, variance of $e^{-bB \cdot C}$ for various values of ψ and showed them in Fig.3-3(d) in blue, yellow and black respectively . When ψ varies from 0 to $\pi/2$, $B \cdot C$ changes from a typical one fiber diffusion to a two fiber diffusion, and from $\pi/2$ to π $B \cdot C$ changes back to the same shape as $\psi = 0$. The graph of CRE shows the value of CRE decreases when $B \cdot C$ varies from one-fiber diffusion to two-fiber diffusion, and increases when $B \cdot C$ gradually changes from two-fiber diffusion backs to one-fiber diffusion.

R_2 cannot detect multi-fiber diffusion as it measures the significance of the second order components in SHS. Nonsignificant difference between R_2 and FA is observed from the images in Fig.3-4(a) and (b). But CRE differs much from R_2 and FA. In Fig.3-3(d) the graph CRE is much steeper than the others. In Fig.3-4, visually, contrast of CRE is much better than that of FA and R_2. Furthermore, the smallness of magnitude of R_2 or FA is unable to distinguish between isotropic and two-fiber diffusion, while that of CRE does better job. Note, CRE is comparable to FA or RA_2 in detecting Gaussian diffusion.

Next we discuss from the theoretical point of view why CRE beats variance in characterizing diffusion anisotropy. Let X be a random variable, $Var(X)$ be its variance. According to proof in [94], $E(|X - E(X)|) \leq 2CRE(X)$. In our case, X is e^{-bd} whose magnitude is multiple of $10^{-2} < 1$, so we have $Var(X) = E(|X - E(X)|^2) \leq E(|X - E(X)|) \leq 2CRE(X)$. Our experiment results show that magnitude of CRE is almost 10 times of that of $Var(X)$. Higher magnitude of CRE makes it less sensitive to rounding errors. One more result from [94] is $CRE(X) \leq \sqrt{Var(X)}$, thus $Var(X) \leq 2CRE(X) \leq 2\sqrt{Var(X)}$, which implies CRE has smaller range than variance does. This is verified by Fig.3-3(d) where the graph of variance is way below that of CRE. Moreover, in Fig.3-4(d), which representing the the variance of e^{-bd}, the Genu/Splenium of corpus callosum is so bright that regions besides it are not clearly visualized , so CRE is much better than variance visually.



(a) (b) (c) (d)

Figure 3-5. (a)-(b). Characterization: black, gray, and white regions represent the voxels with isotropic, one-fiber, and two-fiber diffusion, respectively.(a) using 55 measurements, (b) using 12 carefully selected measurements. (c) Image of CRE calculated from 12 measurements. (d) Characterization results of the region inside the red box in (a) using CRE (top) and variance (bottom) based on 55 measurements. Red arrows point to a voxel that is wrongly characterized as one-fiber diffusion by using variance but correctly classified as two-fiber diffusion using CRE.

Fig.3-5(a) shows a partition of isotropic,one-fiber and two-fiber diffusion based on ADC calculated from 55 measurements. The black, gray, white voxels are identified as isotropic, one-fiber and two-fiber diffusion, respectively. The characterization is consistent

with the known fiber anatomy. Fig.3-5(b) represents the characterization result based on the ADC estimated from 12 measurements. It is very close to that from 55 measurements. CRE based on ADC estimated from 12 measurements (Fig.3-5(c)) is also comparable to that from 55 measurements (Fig.3-4(c)). Thus our characterization is not sensitive to number of measurements. Fig.3-5(d) illustrates a two-fiber diffusion voxel(pointed by red arrow) that is incorrectly characterized as one-fiber diffusion using variance (bottom image) but characterized as two-fiber correctly using CRE (top image). This further verify superiority of CRE over variance in characterizing diffusion anisotropy.

3.3 Summary

In this chapter, we present a novel variational framework for simultaneous smoothing and estimation of ADC profiles depicted by two diffusion tensors. To our knowledge this is the first attempt to use the least amount of measurement to detect two-fiber diffusion from human brain HARD MRI data. We also demonstrated our algorithm for using CRE of e^{-bd} to characterize the diffusion anisotropy.

Our experiments on two sets of human brain HARD MRI data showed the effectiveness and robustness of the proposed model in the estimation of ADC profiles and the enhancement of the characterization of diffusion anisotropy. The characterization of diffusion from the proposed method is consistent with the known neuroanatomy.

In this article, we have not included the work for determination of fiber directions and the method for automated fiber tracking. The study addressing these problems will be reported in separate papers.

CHAPTER 4
RECONSTRUCTION OF INTRA-VOXEL STRUCTURE FROM HARD WEIGHTED IMAGE

4.1 Determination Of Fiber Directions

In our approach of determining fiber directions, the first step is to recover the ADC profiles d from noisy HARD data by using model (2–9) and (2–10). Then, from the SH representation of the recovered d we define

$$R_0(\mathbf{x}) = \frac{|A_{0,0}(\mathbf{x})|}{A(\mathbf{x})}, \; R_2(\mathbf{x}) = \frac{\sum_{m=-2}^{m=2} |A_{2,m}(\mathbf{x})|}{A(\mathbf{x})},$$

where $A(\mathbf{x}) = \sum_{l=0,2,4} \sum_{m=-l}^{l} |A_{l,m}|(\mathbf{x})$. The voxels with significant R_0 and R_2 are identified as *strong* isotropic diffusion and one-fiber diffusion, respectively. The union of all these voxels are denoted by Ω_1. On Ω_1 $f_1$in (1–4) should be close to 1.

Under the assumption of $p_t(\mathbf{r})$ being a mixture of two Gaussians, the diffusion is modelled by (1–4) with $n = 2$. The combination of (1–4) with $n = 2$ and (1–6) yields

$$e^{-bd(\mathbf{x},\theta,\phi)} = \sum_{i=1}^{2} f_i e^{-b\mathbf{u}^T D_i(\mathbf{x})\mathbf{u}}, \tag{4–1}$$

where $\mathbf{u}^T = (sin\theta cos\phi, sin\theta sin\phi, cos\theta)$. To estimate D_i and f_i in (1–5) we minimize the following function:

$$\min_{L_1,L_2,f} \int_\Omega (\sum_{i=1}^{2} |\nabla L_i|^{P_i(\mathbf{x})} + |\nabla f|^{P_f(\mathbf{x})}) d\mathbf{x} + \lambda_1 \int_{\Omega_1} (f_1 - 1)^2 d\mathbf{x}$$

$$+ \lambda_2 \int_\Omega \int_0^{2\pi} \int_0^{\pi} |\sum_{i=1}^{2} f_i e^{-b\mathbf{u}^T L_i L_i^T \mathbf{u}} - e^{-bd}|^2 sin\theta d\theta d\phi d\mathbf{x}, \tag{4–2}$$

with the constraint $L_i^{m,m} > 0$. In (4–2) for $i = 1, 2$ $\lambda_i > 0$ is a parameter, $p_i(\mathbf{x}) = 1 + \frac{1}{1+k|\nabla G_\sigma * \nabla L_i|^2}$, $p_f(\mathbf{x}) = 1 + \frac{1}{1+k|\nabla G_\sigma * \nabla f|^2}$, L_i is a lower triangular matrix such that $D_i = L_i L_i^T$, that is the Cholesky factorization for D_i to achieve the positive definite constraint on D_i (see [32]). $|\nabla L_i|^p = \sum_{1 \le m,n \le 3} |\nabla L_i^{m,n}|^p$;

The first two terms in (4–2) are the regularization terms By the choice of $p_i(\mathbf{x})$ (similarly for p_f), in the homogeneous region image gradients are close to zero and $p_i(\mathbf{x}) \approx 2$, the smoothing is isotropic. Along the edges, image gradient makes $p_i(\mathbf{x}) \approx 1$, the smoothing is the total variation based and only along the edges. At all other locations, the image gradient forces $1 < p < 2$, and the diffusion is between isotropic and total variation based, and varies depending on the local properties of the image. Therefore, the smoothing governed by this model well preserves relevant features in these images. The third term in (4–2) is forcing $f \approx 1$ on Ω_1. The last term is the nonlinear data fidelity term based on (4–1).

The fiber orientations at each voxel are determined by the directions of the principle eigenvectors of D_1 and D_2. For the voxels where f (or $1 - f$) is significantly large, we consider $1 - f$ (or f) as zero, and (1–5) with $n = 2$ reduces to the Gaussian diffusion model.

4.2 Experimental Results

We applied model (4–2) to a set of HARD MRI human data. The raw HARD MR images were obtained using a single shot spin-echo EPI sequence. The imaging parameters for the DW-MRI acquisition are repetition time $(TR) = 1000ms$, $echotime(TE) = 85ms$. Diffusion-sensitizing gradient encoding is applied in fifty-five directions with $b = 1000s/mm^2$. Thus, a total of fifty-six DW images with the matrix size $=256 \times 256$ were obtained for each slice, and images through the entire brain are obtained by 24 slices. The slice is transversally oriented and the thickness is $3.8mm$, and intersection gap between two contiguous slices is $1.2mm$. The field of view (FOV) $=220mm \times 220mm$.

The figures below show our experimental results from a particular subject in a brain slice through the external capsule. In this experiment we first recovered the ADC profiles d using (2–9)-(2–10), and defined Ω_1 as the set of the voxels, where $R_0 > 0.8416$ or $R_2 > 0.1823$. These thresholds were selected using the histograms of R_0 and R_2. Then, we

solved the minimization problem (4–2) by the energy decent method. The information of $f \approx 1$ on Ω_1 was also incorporated into the selection of the initial f.

By solving (4–2) we obtained the solutions L_i and f, and consequently, $D_i = L_i L_i^T$ $(i = 1, 2)$. Fig.4-1a represents the model solution f. Function $f \approx 1$ on the dark red regions. The voxels in these regions are identified as isotropic or one-fiber diffusion. This is consistent to the known neuroanatomy. Figs. 1c and 1d show the color representation of the directions of the principle eigenvectors for $D_1(\mathbf{x})$ and $D_2(\mathbf{x})$, respectively. By comparing the color-coding in Figs.4-1c and 1d with the color pie shown in Fig.4-1b, the fiber directions are uniquely determined. The representation in Fig.4-1b is implemented by relating the azimuthal angle (ϕ) of the vector to color hue (H) and the polar angle ($\theta \geq \pi/2$) to the color saturation (S). Slightly different from [95], we define $H = \phi/2\pi$, $S = 2(\pi - \theta)/\pi$, and $Value(V) = 1$ in SHV. If the direction of the principle eigenvector is represented by (ϕ, θ), the fiber orientation can be described by either (θ, ϕ) or $(\pi - \theta, \phi + \pi)$. We express the vectors in the lower hemisphere, i.e. $\theta \geq \pi/2$. The upper hemisphere is just an antipodally symmetric copy of the lower one. The xy plane is the plane of discontinuity.

Fig.4-2 shows the shapes of $d(\mathbf{x}, \theta, \phi)$ together with the fiber directions at 4 particular voxels. The blue and red arrows indicate the orientations of the fibers determined from the principle eigenvectors of D_1 and D_2 respectively. The last shape corresponds to isotropic diffusion. Figs.4-1 and Fig.4-2 indicate that our model (4–2) is effective in recovering the intra-voxel structure.

To examine the accuracy of the model in recovering fiber directions, we selected a region where the diffusion is known as one-fiber diffusion inside the corpus callosum. For each voxel in this region we computed the direction in which d is maximized. This direction vector field is shown in Fig.4-3a. On the other hand we solved (4–2) and obtained the model solution $f \approx 1$ on this region. The direction field generated from the

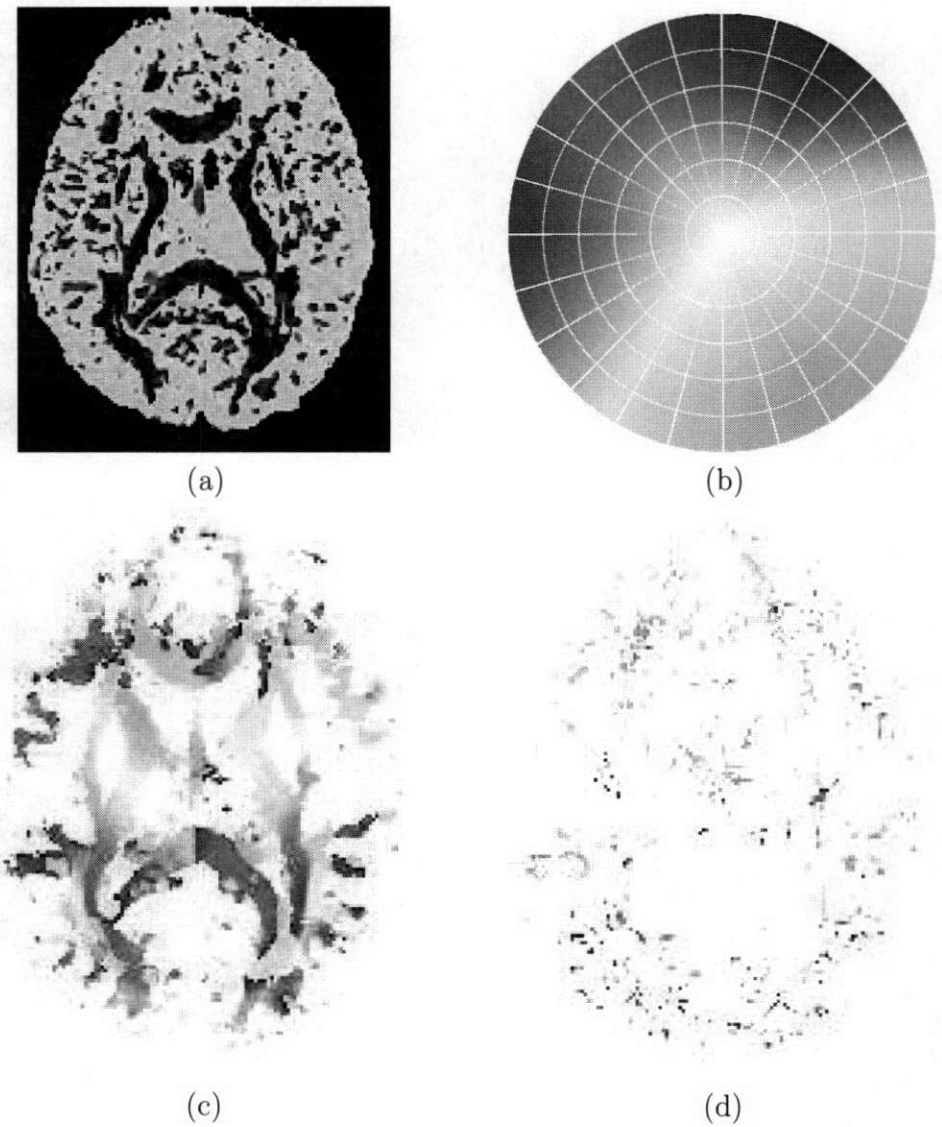


Figure 4-1. (a). Model solution f, (b). color pie, (c). color-coding of the 1st fiber direction mapping, (d). color-coding of the 2nd fiber direction mapping.

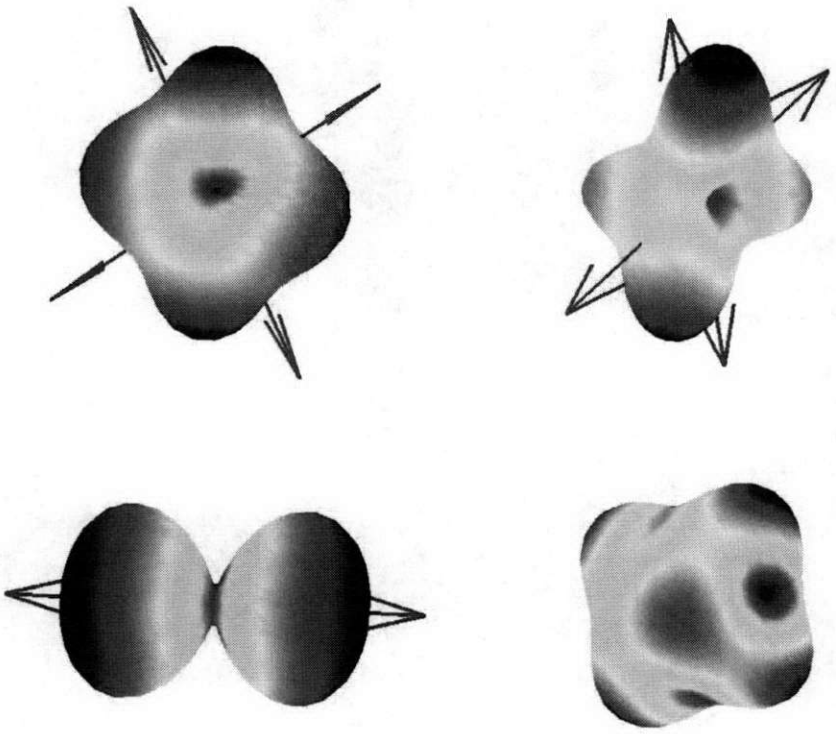


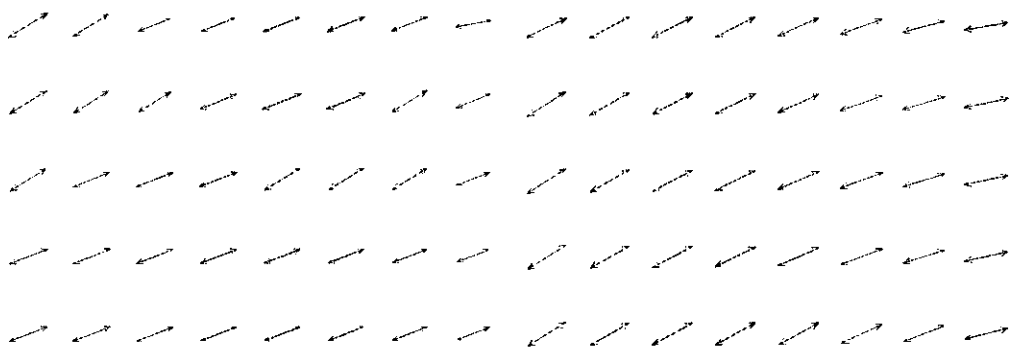
Figure 4-2. Shape of d, and the orientations of the principle eigenvectors of D_1 (blue) and D_2 (red) at 4 particular voxels

principle eigenvector of D_1 was then shown in Fig.4-3b, in which the vector field is not only well preserved but also more regularized due to the regularization terms in the model.

4.3 Conclusion

A novel model for determination of fiber directions for biGaussion diffusion from HARD MRI is presented. In this model two tensor fields are recovered by simultaneous smoothing and data fitting, as well as incorporating the information on voxel classification of diffusion anisotropy. The numerical results indicate the effectiveness of the model in recovering intra-voxel structure.

The choice of the parameters in (4–2) and the determination of the region Ω_1 would influence the results. Our choice is made based on the principle that the one-fiber direction from the model agrees with the direction in which d is maximized. We will



(a) (b)

Figure 4-3. Fiber direction field obtained by (a) maximizing d, (b). the principle eigenvector of D (solution of (4–2))

study in the future how to increase the accuracy in the estimation of the orientations of crossing fibers within a voxel.

CHAPTER 5
RECONSTRUCT WHITE MATTER FIBER TRACES USING MULTI-TENSOR DEFLECTION IN DWI

5.1 Recovery Of Multi-Tensor Field In HARD MRI

We assume the data acquisition noise is additive and the distribution is modeled as a mixture of two Gaussians(1–12). The goal of this section is to recover a smooth multi-tensor field from the noisy HARD MRI data. By Cholesky factorization theorem, a symmetric matrix D is positive definite if and only if $D = LL^T$, where L is a lower triangular matrix. Uniqueness of the factorization is ensured by positiveness of the diagonal entries of L. To guarantee the positive definiteness of diffusion tensor D_1, D_2, let $D_i = L_i L_i^T$, for $i = 1, 2$, L_i is a lower triangular matrix with positive diagonal entries. Constraint $0 \leq f \leq 1$ is fulfilled through variable relaxation method. For example, we let $f(\mathbf{x}) = .5 + \frac{\arctan(\omega(\mathbf{x}))}{\pi}$ which is an increasing smooth function of ω, it is obvious that f defined in this way lies in $[0, 1]$. To get a regularized estimation of functions $f(\mathbf{x}), D_1(\mathbf{x}), D_2(\mathbf{x})$, we solve the following constrained minimization problem:

$$\min_{L_1(\mathbf{x}),L_2(\mathbf{x}),f(\mathbf{x})} \int_\Omega (\sum_{i=1}^{2}\sum_{m=1}^{3}\sum_{n=1}^{m} \alpha|\nabla L_i^{mn}(\mathbf{x})| + \beta|\nabla f(\mathbf{x})|)d\mathbf{x} + \tag{5–1}$$
$$\int_\Omega \int_0^{2\pi} \int_0^{\pi} |s_0(\mathbf{x})(fe^{-b\mathbf{u}^T L_1 L_1^T \mathbf{u}} + (1-f)e^{-b\mathbf{u}^T L_2 L_2^T \mathbf{u}}) - s(\mathbf{x},\theta,\phi)|^2 sin\theta d\theta d\phi d\mathbf{x}$$

with the constraint $L_i^{mm} > 0$, for $i = 1, 2, m = 1, 2, 3$. Where we denote the mn^{th} entry of L_i by L_i^{mn} , α, β are weighting factors. $\mathbf{u} = (sin\theta cos\phi, sin\theta sin\phi, cos\theta)^T$ with $0 \leq \theta < \pi$ and $0 \leq \phi < 2\pi$. The first two terms are the regularization terms, without which model(5–1) will be an ill-posed problem and the last term is the nonlinear data fidelity term based on (1–12).

We employed a gradient descent scheme to solve the minimization problem(5–1). Initials of f, L_1, L_2 are carefully chosen to avoid sticking on local minimum. For conciseness, Euler-Lagrangian equations of f, L_1^{11} only are shown as followings:

$$\begin{aligned}\frac{\partial f}{\partial t} &= -2s_0(\mathbf{x})\int_0^{2\pi}\int_0^{\pi}[s_0(\mathbf{x})(fe^{-b\mathbf{u}^T L_1 L_1^T \mathbf{u}} + (1-f)e^{-b\mathbf{u}^T L_2 L_2^T \mathbf{u}}) - s(\mathbf{x},\theta,\phi)]\cdot \\ &\quad (e^{-b\mathbf{u}^T L_1 L_1^T \mathbf{u}} - e^{-b\mathbf{u}^T L_2 L_2^T \mathbf{u}})sin(\theta)d\theta d\varphi + \beta div(\frac{\nabla f}{|\nabla f|}) \qquad (5\text{–}2)\\ \frac{\partial L_1^{11}}{\partial t} &= 4bfs_0(\mathbf{x})\int_0^{2\pi}\int_0^{\pi}[s_0(\mathbf{x})(fe^{-b\mathbf{u}^T L_1 L_1^T \mathbf{u}} + (1-f)e^{-b\mathbf{u}^T L_2 L_2^T \mathbf{u}}) - s(\mathbf{x},\theta,\phi)]\cdot \\ &\quad e^{-b\mathbf{u}^T L_1^T \mathbf{u}}(\mathbf{u_1}L_1^{11} + \mathbf{u_1}\mathbf{u_2}L_1^{21} + \mathbf{u_1}\mathbf{u_3}L_1^{31})sin(\theta)d\theta d\varphi + \alpha div(\frac{\nabla L_1^{11}}{|\nabla L_1^{11}|}) \qquad (5\text{–}3)\end{aligned}$$

5.2 White Matter Fiber Tractography

Results of the above section provide a smooth multi-tensor vector field and a smooth volume fraction field f, fiber tractography based on which is almost not sensitive to noise and thus more accurate. In this section, we will provide an improved line propagation algorithm for reconstructing white matter fiber traces. Line propagation scheme is defined by: $\mathbf{x}(t+1) = \mathbf{x}(t) + \mathbf{v}(t+1)\delta$, where $\mathbf{x}(t)$ is the position vector in R^3 of the streamline at time t, $\mathbf{v}(t+1)$ is a unit vector that leads the particle from the position $\mathbf{x}(t)$ to the next step $\mathbf{x}(t+1)$, and δ is the step size.

In DTI data, Westin et.al [41] used the entire tensor D at location $\mathbf{x}(t)$ to determine $\mathbf{v}(t+1)$ as $D\cdot\mathbf{v}(t)$. They also provided a less sensible scheme which dynamically modulates PE $\mathbf{e_1}$ of tensor D and tensor deflection $D\cdot\mathbf{v}(t)$ contributions to trace steering:

$$\mathbf{v}(t+1) = \alpha\mathbf{e_1} + (1-\alpha)((1-\beta)\mathbf{v}(t) + \beta D\cdot\mathbf{v}(t)) \qquad (5\text{–}4)$$

Where α and β are user-defined weighting factors that vary between 0 and 1, $\mathbf{e_1}$ and $\mathbf{v}(t)$ are normalized before used. Since PE doesn't make sense at voxels with degenerated

anisotropy, we define $\mathbf{v}(t+1)$ as

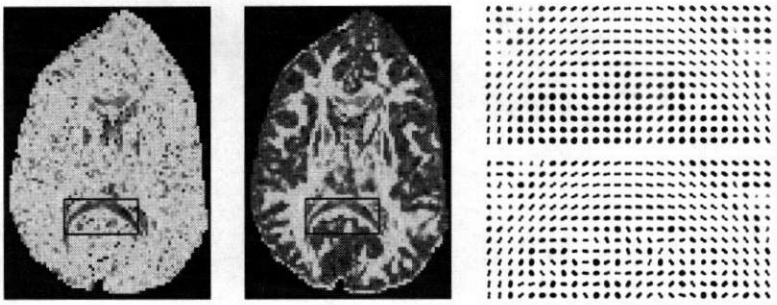
$$\mathbf{v}(t+1) = (1-\beta)\mathbf{v}(t) + \beta D \cdot \mathbf{v}(t) \tag{5–5}$$

Here $D \cdot \mathbf{v}(t)$ is also normalized before used. Normalization of it was ignored in [41], but this is essentially necessary as for human brain HARD MRI data, norm of $D \cdot \mathbf{v}(t)$ is usually in the order of 10^{-3}, tensor deflection without normalization would not contribute as much as expected. To make following explanation concise, let $f_1 = f, f_2 = 1 - f$. At each voxel, for $i = 1, 2$, apply (5–5) to D_i to get $\mathbf{v}_i(t+1)$. Define corresponding step size as $\delta_i = c f_i(\mathbf{x}(t)) FA_i(\mathbf{x}(t)) \mathbf{v}(t) \cdot \mathbf{v}_i(t+1)$ with c a fixed constant, $FA_i(\mathbf{x}(t))$ the fractional anisotropy(FA) corresponding to tensor D_i. As we know, if f_i is very close to 0, channel D_i could be ignored; if FA_i is very low, anisotropy of D_i is low; if $\mathbf{v}(t) \cdot \mathbf{v}_i(t+1)$ is low, there is too much bending between $\mathbf{v}(t)$ and $\mathbf{v}_i(t+1)$. So fiber tracking should be terminated at channel D_i when anyone of the above quantities is low. This could simply be done by setting a threshold to step size so that channels with step size less than this threshold are terminated. The threshold is a statistical value obtained through a large size of experiments. This self adapting step size constrains propagation speed in regions with high curvature and low diffusion anisotropy while increases speed in regions with low curvature and high diffusion anisotropy, it also automatically terminates fiber tracking at channel(s) with extremely low step size(s).

Our scheme generalizes the tensor line propagation algorithm. We call it multi-tensor line propagation(MTEND). The challenging aspect of this method is the estimation of $\mathbf{v}(t+1)$'s at non-grid points. We linearly interpolate f and 6 entries of D_1, D_2 respectively , $\mathbf{v}(t+1)$'s are then calculated using (5–5) based on the interpolated f, D_i's.

5.3 Experimental Results

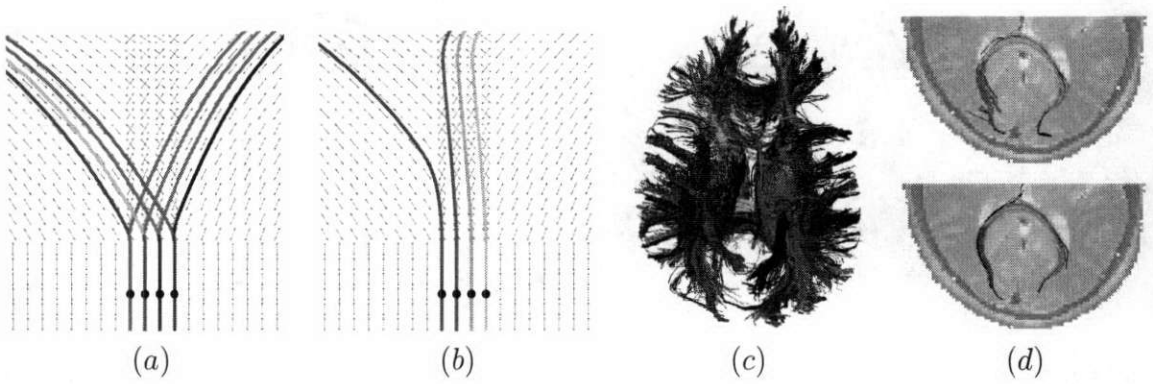
In this section we present synthetic as well as real data experiments. We did experiments on a set of subjects, but only list results of one subject for demonstration.



(a) (b) (c)

Figure 5-1. (a)-(b) FA maps of the first tensor field D_1 obtained using Parker et. al's([17]) and proposed model(5–1) respectively. (c)Images of ADC profile $\mathbf{u}^T D_1 \mathbf{u}$ of voxels inside the selected regions in (a)and (b), the top one and the bottom one are obtained using D_1 calculated using proposed model(5–1) and Parker's method respectively.

The first experiment is to demonstrate the superiority of the proposed model (5–1) over Parker et. al's method([17]) in recovering smooth multi-tensor field as well as the volume fraction f using human brain HARD MRI data. The data set consists of 33 diffusion weighted images as well as one image in the absence of a diffusion-sensitizing field gradient. 27 evenly spaced axial planes with 128×128 voxels in each slice are obtained using a 3T MRI scanner with a single shot spin-echo EPI sequence. Slice thickness is $3.8mm$, gap is 0 between two consecutive slices , repetition time $(TR) = 1000ms$, $echotime(TE) = 85ms$ and $b = 1000s/mm^2$, and the field of view (FOV) = $200mm \times 200mm$. Fig.5-1$(a) - (c)$ compare maps of FA of the first tensor field D_1 and apparent diffusion coefficient(ADC) profiles of D_1, i.e. $\mathbf{u}^T D_1 \mathbf{u}$ using solutions obtained from Parker et.al's([17]) and proposed model(5–1). It is obvious that all the results obtained from our model are much more smooth and more reasonable than that got from Parker's method. Specifically, Fig.5-1(b)(result of proposed model(5–1)) gives a reasonable FA image from which we are able to distinguish voxels with high anisotropy(red region) from that with low anisotropy(blue region). While in Fig.5-1(a)(result of Parker's method) FA collapses except in regions around the corpus callosum. In Fig.5-1(c), at each voxel we show shape of ADC profile $\mathbf{u}^T D_1 \mathbf{u}$ in 41×21 directions, i.e. $41 \times 21\mathbf{u}$'s . In the top image of Fig.5-1(c)(result of proposed model(5–1)), shapes of ADC profile change smoothly

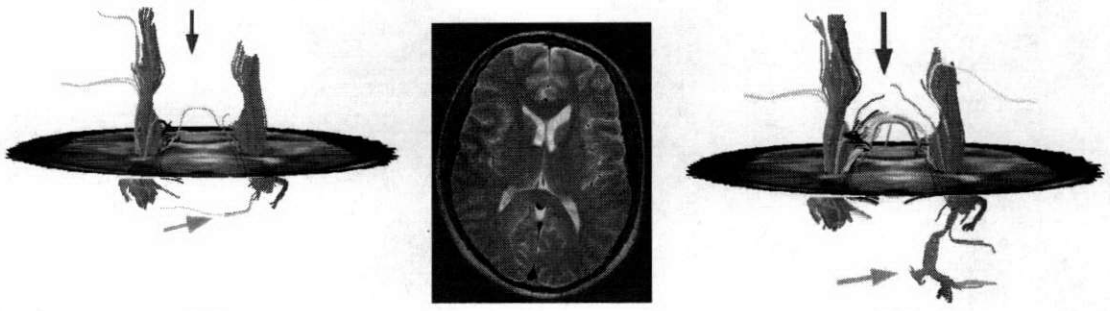


(a) (b) (c) (d)

Figure 5-2. (a) Traces recovered by using MTEND based on simulated multi-tensor field, the black points at the bottom are seeds. (b) Traces recovered by using TEND based on simulated DTI data. (c) Axial view of fiber tracking result of the whole brain by MTEND algorithm. (d) Axial view of fiber tracking results using MTEND(top),TEND(bottom) algorithm .

from voxel to voxel, and in the region below corpus callosum, voxels which are mostly likely to be of isotropic diffusion have sphere-shaped ADC. But in the bottom image of Fig.5-1(c)(result of Parker's method) shapes of ADC profile jump a lot from voxel to voxel, especially in regions outside the corpus callosum.

The second experiment is to show MTEND algorithm outdoes TEND algorithm in reconstructing fiber traces including bifurcation. This is done on simulated data. We first simulate a $20 \times 20 \times 3$ multi-vector field shown as blue arrows in Fig.5-2(a)(b), where voxel with one arrow owns only one tensor that has the direction shown by this blue arrow as the principle eigenvector, while voxel with two arrows owns two tensors that have the directions shown by the two blue arrows as the principle eigenvectors. Secondly, we construct a multi-tensor field so that the multi-vector field is the corresponding principle eigenvector field. Raw DTI data are finally simulated based on the simulated multi-tensor field using(1–12) with $s_0 = 400, b = 1000$, $f = 1$ at voxels with one vector,$f = .5$ at voxels with two vectors, and 6 $\mathbf{u}$'s which are uniformly distributed on a sphere. Apply MTEND algorithm to the multi-tensor field while apply TEND algorithm to the single-tensor field obtained from DTI data, we get results shown in Fig.5-2(a),(b) respectively. The four black points are the initiations, i.e. seeds of the fiber tracking. Nice bifurcations are



(a) (b) (c)

Figure 5-3. (a)(c) are tracking results using TEND, MTEND method respectively. (b) Anatomic image of one slice with red regions as the seeds of tracking.

observed in Fig.5-2(a), and they take place in voxels with 2 diffusion tensors as expected. In comparison, no bifurcation is visualized from Fig.5-2(b) and only the most left fiber trace goes almost along the vector field, while the other 3 fiber traces do not make sense at all. This verified the efficiency of MTEND algorithm based on multi-tensor field outdoes TEND based on single-tensor field in recovering fibers with branching.

Next, experiments on human brain HARD MRI data are shown. The main aim is to show that MTEND and TEND work similarly in the corpus callosum region where Gaussian diffusion is dominant. But they differ in regions with non-Gaussian diffusion. Before comparing them, we show in Fig.5-2(c) an axial view of the whole brain's fiber traces . The initiations of tracking are set at all the anisotropic voxels in the whole brain volume. Different strong fiber bundles and branches are clearly visualized. They are consistent with known neuroanatomy.

Fig.5-2(d) shows axial view of tracking results around the corpus callosum region using MTEND(top),TEND(bottom) algorithm. Tracking results are embedded on a 2D anatomic image. Tracking start from a small portion inside the corpus callosum. No significant difference is observed as in corpus callosum region, Gaussian distribution is dominant, biGuassian model with $f \simeq 1$ and Gaussian model work equivalently in recovering a single-tensor field.

Finally, we select two regions of interest(ROI) from the internal capsule(red region in Fig.5-3(b)) for another set of comparison. We set all the anisotropic voxels in the whole brain volume as seeds, then apply MTEND based on multi-tensor field recovered using model(5–1) and TEND based on single-tensor field recovered using Gaussian diffusion model to reconstruct fiber traces separately. In MTEND we set $\beta = 0.9$, threshold value of step size δ to be 0.1 which is obtained from a large size of experiments. Only those fibers passing through the ROIs are retained and shown in Fig.5-3(a)(c) for TEND and MTEND respectively. Clearly, MTEND method recovers more branching fibers than TEND method does. Specifically, it happens in 3 different locations: one is at the lower right position and directed by orange arrows. Bunches of fiber traces with several branches are nicely shown up in Fig.5-3(c), but they do not appear in Fig.5-3(a). The second one is located at the middle and directed by blue arrows. A strong bundle connecting the left portion and the right portion is clearly visualized in Fig.5-3(c) but only one fiber trace is shown in Fig.5-3(a). The third one lies in the most upper left position: Fig.5-3(c) looks thicker and includes more fibers in each branching than Fig.5-3(a) does. The main reason for the difference is that voxels involving branching in MTEND method are characterized as isotropic, so TEND algorithm terminates at these voxels.

5.4 Conclusion

A new variational framework for simultaneous reconstruction and regularization of multi-diffusion tensor field and a new fiber tractography algorithm based on multi-tensor field are provided. The performance of the proposed model has been evaluated on synthetic data and several human subjects' HARD MR images. The experimental results indicate that proposed model(5–1) for recovering multi-tensor field together with MTEND for reconstruction of white matter fiber traces work more accurately than Gaussian diffusion model together with TEND.

The proposed model is under the assumption that the probability density function of diffusion is of linear combination of two Gaussians. This results in 13 unknowns at each

voxel, and hence at least 13 diffusion weighted images acquisition is required to solved out 13 unknowns accurately. Model that does not require specific assumption on diffusion and that requires less diffusion weighted images will be addressed in separate papers.

CHAPTER 6
DEFORMABLE MEDICAL IMAGE REGISTRATION

As mentioned in section (1.3), one promising approach for deformable registration is finding an optimal deformation field $U(X)$ by minimizing an energy functional that consists of a similarity measure between the reference (target) image and transformed template (source image) and a smoothness measure for the deformation field $U(X)$. $\|\nabla U\|_{L^2}$ can be chosen as the regularity term, but for cases whether image pairs are in same modality or in different modalities, the similarity measures can vary a lot. Here we just talk about them separately.

6.1 Same-Modality Deformable Image Registration

Let two images $S(x)$ and $T(x)$ be given as the template and the reference images defined on $\Omega \subset R^n$, respectively. To find an optimal deformation field $u(x)$ (displacement), Lu et al. [58] minimizes the following energy

$$E(u) := \frac{1}{2}\int_{\Omega}\{\lambda|\nabla u|^2 + |S(x+u(x)) - T(x)|^2\}dx, \tag{6–1}$$

where λ is a parameter balancing the smoothness of the transformation versus the similarity of the images. The existence of minimizer for (6–1) is proved in appendix (A). The minimization problem was solved by using energy decent method. That is iteratively solving the PDE:

$$\partial_t u(x,t) = \lambda\Delta u(x,t) + (T(x) - S(x+u(x,t)))\nabla S(x+u(x,t)) \tag{6–2}$$

Their numerical results showed the effectiveness of this model in radiotherapy planning and evaluation that incorporates internal organ deformation.

However, the parameter λ in this model effects registration results greatly, and the selection of λ is always a difficult problem.

One commonly used method is to set λ as a constant, the resulting deformation field is then usually sensitive to λ: if λ is too small, the deformation will be not smooth; if λ is

too large, algorithm will converge slowly and regions involving large deformation can not be matched well in short time. Since λ is a universal weight parameter, there is tradeoff between the regions needing large and less deformation.

Intuitively, we wish λ is larger in the beginning of the iteration and reduced as the alignment is getting better. Moreover the dissimilarity measure in (6–1) is sensitive to the presence of more noisy and outliers [46]. To deal with these limitations we propose a simple modification of this model that accommodates local degrees of variability for the matching. Our idea is to consider $S(x + u(x)) - T(x)$ as independent normal random variables indexed by the pixel x with probability density functions

$$p(S(x + u(x)) - T(x)) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{|S(x+u(x))-T(x)|^2}{2\sigma^2}}.$$

To have a better registration between S and T, we replace minimizing $\int_\Omega |S(x + u(x)) - T(x)|^2 dx$ (sum of square distances) by maximizing the log-likelihood function, i.e. finding the optimal deformation field and variance by minimizing the following functional with respect to $u(x)$ and σ:

$$-\int_\Omega \ln p(S(x + u(x)) - T(x)) = \int_\Omega \frac{|S(x + u(x)) - T(x)|^2}{2\sigma^2} + \ln \sqrt{2\pi}\sigma dx.$$

Together with the smoothness constraint for $u(x)$ our model reads as

$$\min_{u(x),\sigma} E(u(x), \sigma) := \min_{u(x),\sigma} \frac{1}{2} \int_\Omega \{\lambda |\nabla u|^2 + \frac{1}{2\sigma^2} |S(x + u(x)) - T(x)|^2\} dx + |\Omega| \ln \sigma \qquad (6\text{–}3)$$

Comparable settings appeared in [96–98] regarding optical flow estimation and in [99] regarding deformable neuroanatomy textbook registration. Proof of existence of a minimizer for (6–3) is in appendix B. By calculus of variation, we have the Euler-Lagrange (EL) equations for (6–3):

$$-\lambda \Delta u(x) + \frac{1}{2\sigma^2}(S(x + u(x)) - T(x)) \nabla S(x + u(x)) = 0 \qquad (6\text{–}4)$$

$$\sigma^2 = \frac{1}{|\Omega|} \int_\Omega |S(x + u(x)) - T(x)|^2 dx. \qquad (6\text{–}5)$$

Inserting (6–5) to (6–4) we get the equation of $u(x)$:

$$-\lambda\Delta u(x,t) + \frac{|\Omega|}{2\int_\Omega |S(x+u(x)) - T(x)|^2 dx}(S(x+u(x)) - T(x))\nabla S(x+u(x)) = 0.$$

This problem is solved by finding the steady state solution of the following problem:

$$\partial_t u(x,t) = \lambda\Delta u(x,t) + \frac{|\Omega|}{2\int_\Omega |S(x+u(x,t)) - T(x)|^2 dx}(T(x) - S(x+u(x,t))\nabla S(x+u(x,t)), \tag{6–6}$$

with the boundary and initial conditions:

$$\frac{\partial u}{\partial n} = 0, \quad on \quad \partial\Omega \times R^+, \qquad u(x,0) = 0, \quad on \quad \Omega.$$

Our model (6–3) differs from Lu et al.'s model (6–2) in the aspect that it allows the deformed template having a variance from the reference. This is especially good for aligning two images whose intensity are not exactly equal or linearly related. Moreover, the weight between the smoothness and similarity in (6–2) is fixed in all integrations, while this weight in (6–6) varies at each iteration depending on the standard deviation of $S(x+u(x)) - T(x)$. The weight on the similarity measure increases when the standard deviation of $S(x+u(x)) - T(x)$ decreases.

6.2 Multi-Modality Deformable Image Registration

For multi-modality image registration more complex intensity similarity measures are necessary. We adopt mutual information as a similarity measure, which has been proven effective for multi-modality image registration [52, 100–102]).

Our approach is replacing the SSD measure in (6–1) by the mutual information (MI) between $S(x+u(x))$ and $T(x)$.

Mutual information between two random vectors X and Y is defined as

$$MI(X,Y) = H(X) + H(Y) - H(X,Y),$$

where

$$H(Z) = -\int_{R^N} p_Z(z) \log p_Z(z) dz,$$

is the Shannon entropy of a random N-vector Z with probability density function (pdf) $p_Z(z)$, and

$$H(X,Y) = -\int_{R^N}\int_{R^N} p_{X,Y}(x,y) \log p_{X,Y}(x,y) dxdy,$$

is the joint entropy of X and Y with the joint pdf $p_{X,Y}(x,y)$.

Consider an image as independent random variables indexed by the voxel x. The mutual information of two images $S(x+u(x))$ and $T(x)$ can be computed as follows: Let

$$f(x) = S(x+u(x)), \quad g(x) = T(x).$$

$$MI(S(x+u(x)), T(x)) =: \int_{R^2} p_{f,g}(i_1,i_2) \log \frac{p_{f,g}(i_1,i_2)}{p_f(i_1)p_g(i_2)} di_1 di_2,$$

where $p_{f,g}(i_1,i_2)$, $p_f(i_1)$ and $p_g(i_2)$ are the joint pdf of f and g, pdf of f, and pdf of g, respectively. One can see from this formula that the mutual information of f and g is the same as the Kullback-Leibler distance of the random variables with the probability density functions $p_{f,g}(i_1,i_2)$ and $p_f(i_1)p_g(i_2)$. Therefore, MI measures mutual dependence of f and g.

Using MI as a similarity measure our model reads as follows:

$$\min_{u(x)} E(u) := \min_{u(x)} \frac{1}{2}\int_\Omega \lambda |\nabla u|^2 dx - \int_{R^2} p_{f,g}(i_1,i_2) \log \frac{p_{f,g}(i_1,i_2)}{p_f(i_1)p_g(i_2)} di_1 di_2. \tag{6-7}$$

Our pdf estimator is based on normalized Gaussian kernel of variance β, i.e.

$$p_{f,g}(i_1,i_2) = \frac{1}{|\Omega|}\int_\Omega G_\beta(f(x)-i_1, g(x)-i_2) dx, \tag{6-8}$$

for some small $\beta > 0$. Then, its marginals are computed by

$$p_f(i_1) = \int_R p_{f,g}(i_1,i_2) di_2, \quad p_g(i_2) = \int_R p_{f,g}(i_1,i_2) di_1.$$

The fact that the first variation of E vanishes at a minimizer yields the EL equation:

$$-\lambda\Delta u(x) + \{\frac{p_f'(i_1)}{p_f(i_1)} - \frac{\partial_1 p_{f,g}(i_1,i_2)}{p_{f,g}(i_1,i_2)}\}\nabla S(x+u(x)) = 0,$$

where ∂_1 denotes the partial derivative with respect to the first intensity variable i_1.

Then, the steady state solution of the following evolution problem gives the minimizer of (6–7).

$$\partial_t u(x,t) = \lambda\Delta u(x,t) + \{\frac{\partial_1 p_{f,g}(i_1,i_2)}{p_{f,g}(i_1,i_2)} - \frac{p_f'(i_1)}{p_f(i_1)}\}\nabla S(x+u(x,t)), \tag{6–9}$$

with the boundary and initial conditions:

$$\frac{\partial u}{\partial n} = 0, \quad on \quad \partial\Omega \times R^+, \qquad u(x,0) = 0, \quad on \quad \Omega.$$

6.3 Numerical Issues

In implementation of (6–6,6–9), Additive Operator Splitting (AOS) scheme, an efficient semi-implicit finite difference scheme is used. Under AOS algorithm, one can make both computational and storage cost be linear in the number of pixels, and can gain an increase of computational efficiency by a factor of 10 under realistic accuracy requirements([89]).

Moreover, a multi-resolution scheme (also called pyramid) is used, in which higher resolution iterations just take the interpolated result from previous resolution as initial. Pyramid scheme can not only help us prevent being trapped in local minima, but also help us get the solution efficiently since getting solution for lower resolution images takes much less time comparing with that of full resolution images, this is essentially helpful for registering large data pairs.

6.4 Experiment Results

This section has two group of examples. One is for same modality image registration and the other is for multimodality case. The first group of four examples are to demonstrated the effectiveness of model (6–3) comparing with model (6–1). The second group of one example is to show the multimodality registration result based on variational method.

6.4.1 Same-modality experiments

For both model (6–3) and (6–1), we use finite difference scheme with steepest gradient descent method. Both iteration schemes have two parts: The first part is Laplacian operator comes from the regularity term, and the second part is deformation force term which comes from the fidelity term of model. If the force term is dominant, then large deformation will be more easy to achieve but the vector field might be not smooth. If the regularity term is dominant, the scheme will generate smooth deformation field but large deformation is hard to get. To get good registration results, we would like to make these two parts reasonably balanced. If we define a ratio of the Fidelity term Over the Regularity term (FOR) as:

For model (6–1)

$$FOR = \frac{\int_\Omega (S(X + U(X)) - T(X))^2 dX}{\int_\Omega |\nabla U(X)|^2 dX}$$

For model (6–3)

$$FOR = \frac{\frac{1}{2\sigma^2}\int_\Omega |S(x + u(x)) - T(x)|^2 dX}{\int_\Omega |\nabla U(X)|^2 dX}$$

We will check the behavior of FOR in experiments.

The first example is to apply both schemes on synthetic images of size 65×65 , here both source and target are smooth images generated by distance function. In the source image I_1, there is a circle with center $(33, 33)$ and radius 30, the intensity is the distance to the circle if a pixel locates inside of the circle and is 0 otherwise, while in the target image I_2, there is a square with center at $(33, 33)$ and with width 61, similarly the intensity is 0 outside of the square and is $30 - d(x)$ inside the square where $d(x)$ is the distance from x to the nearest side. Since the image is very small, we apply our algorithms on the full resolution images directly without using pyramid scheme.

After applying both model (6–3) and model (6–1) on the synthetic images, one can see that the deformed results, either by model (6–3) in Fig 6-1(b) or by model (6–1) in

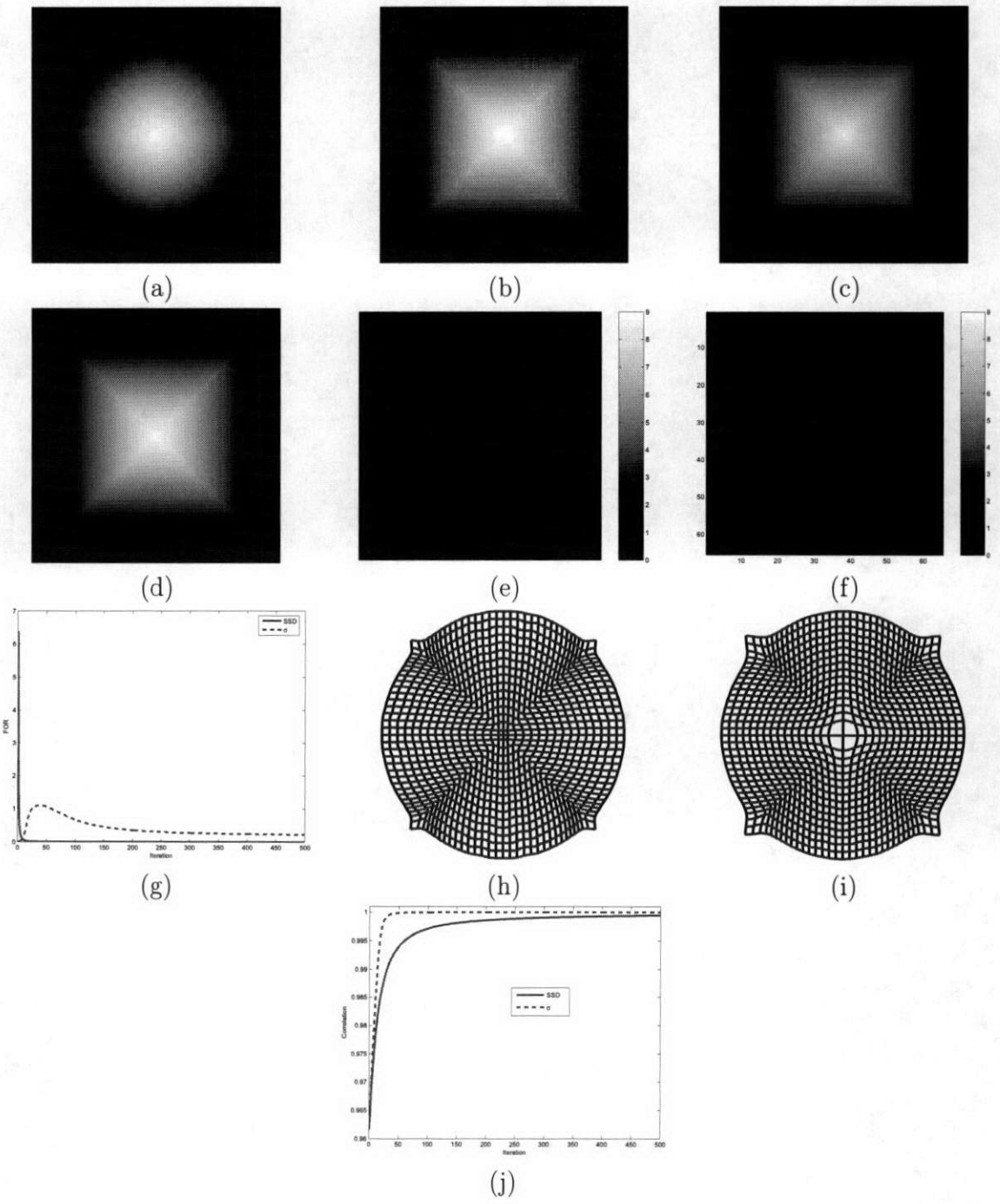


Figure 6-1. Registration for synthetic data. (a,d)Synthetic source and target images respectively; (b,c) registration results by model (6–3) with $= .1, dt = 1.0$ after 500 iterations and model (6–1) with $= .001, dt = 50.0$ after 1000 iterations respectively;(e,f), corresponding difference image;(g) FOR; (h,i) corresponding deformation grids;(j) correlation;

Fig 6-1(c), are close to the target image (d). Fig6-1(e) shows that FOR approaches to 0 rapidly in the iterations of model (6–1)(blue solid line), which makes the force to drive the deformation approach to 0 during iterations, hence scheme (6–2) make the vector field deform slowly after few iterations, while FOR in model (6–3) can keep the two terms in a reasonable ratio (red dashed line), this makes scheme (6–6) converges faster than (6–2).

Both schemes give good deformation grids Fig 6-1(h, i). In this experiment, the ground truth is known: since we are try to align two images based on intensity only, the corresponding intensity level sets in the image pair should be matched, namely, to fit a level curve which form as a closed square in I_2, the corresponding part in I_1 should be the same intensity level set which forms a circle, this is observed in both Fig 6-1(h) and 6-1(i). Deformed grid from (6–3) in 500 iterations is even slightly better than grid got in 1000 iterations from (6–1), this comparison again shows faster convergence of model (6–3).

Fig 6-1(j) shows the cross correlation between deformed image and the target image during first 500 iterations. One can see that CC of scheme (6–6) approaches to 1 in about 50 iterations while scheme (6–2) takes about 500 iterations. Faster convergence of model (6–3) is again concluded here.

In a word, Fig 6-1 shows that both models can get good results if data is good, but model (6–3) converges faster than model (6–1).

Fig 6-2 is to compare performance of schemes on 3D CT lung data. Large deformation is observed from CT lung volumes taken between the status of exhale (Fig 6-2(a)) and the status of inhale (Fig 6-2(b)), which also generate motion of tumor (the isolated white spot in coronal view). Within limited number of iterations, using model (6–1) can only catch a close solution to this large deformation because of the regularity term hinders the large deformation(Fig 6-2(d,f)), another way to say is that the deformation force decrease dramatically because FOR approaches to 0 rapidly in scheme (6–2) , while the model (6–3) can make the large deformation happen while maintain the smooth deformation

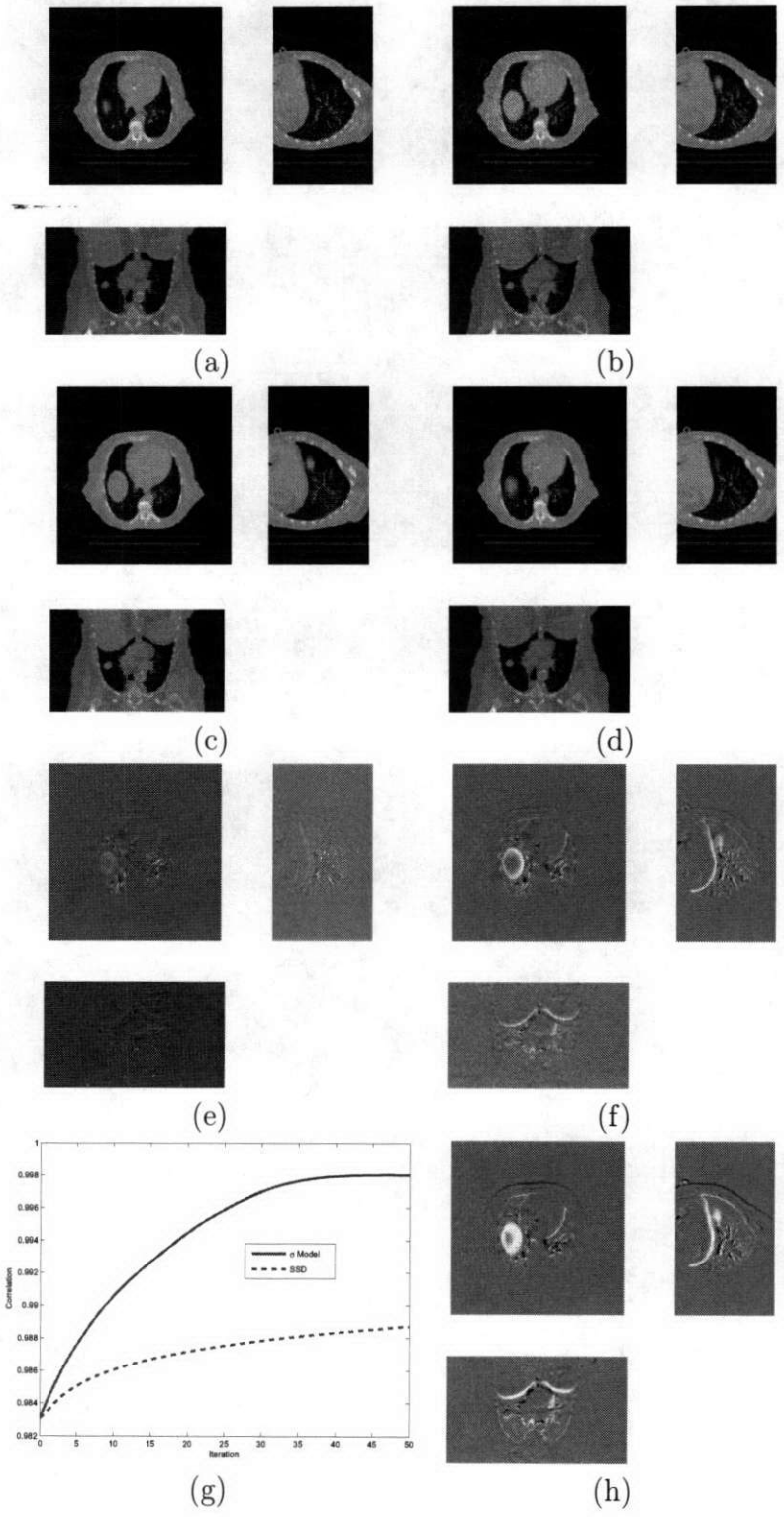


Figure 6-2. Registration for CT Lung data . (a,b)Lung source and target images respectively; (c,d) registration results by model (6–3) and (6–1) respectively; (e,f) volume difference with target by subtracting results by model (6–3) and (6–1) respectively;(g) correlation;(h) Original volume difference.

field since deformation is more balanced by the variance(Fig 6-2(c,f)). Fig 6-2(e) shows, in sense of CC, scheme (6–6) converges faster than scheme (6–2).

Fig 6-3 is to compare performance of schemes on 3D MRI volumes with size 120 × 187 × 51. These two data were collected in two adjacent days, and in the data, large motion is observed between Fig 6-3(a) and (b). After 20 iterations, model (6–1) gives a deformation which is not quit good (d,f,g), while the model (6–3) behaviors better in Fig 6-3(c,e,g). Fig 6-3(g) again shows scheme 6–3 converges faster than scheme 6–2 from the curve of CC. The difference images show the advantages of model (6–3) than model (6–1) by comparing Fig 6-2(e,f,g).

Fig 6-4 is to compare performance of schemes on noised 3D MRI volumes, by adding Gaussian noise with 0 mean and variance .005 to images shown in Fig 6-3(a,b) . After 20 iterations, model (6–1) a deformation which is far from good enough (d,f), while results in 6-4(c,e) shows the model (6–3) behaviors as well as in Fig 6-3(ce). Fig 6-4(g) again shows scheme (6–3) converges faster than scheme (6–2) from the curve of CC. In one word, Fig (6-4) shows that model (6–3) can handle noise much better than model (6–1).

To eliminate the effect of pyramid scheme, here all these four examples are doing iterations on full resolution directly. After combining pyramid scheme and AOS algorithm (which is parallelizable), all these 3D experiments can be done within 10 to 20 seconds. After parallelization, the time can be reduced to seconds, which is very close to clinical applications.

6.4.2 Multi-modality experiments

This section shows the deformable registration results by model (6–7).Since the computation of joint pdf for image pair (6–8) involves exponential function, which is very expensive when the volumes have large sizes. To improve efficiency of the scheme 6–9, both pyramid scheme and AOS algorithm are occupied here.

In Fig 6-5, model (6–7) is applied to register 3D CT volume to 3D MRI volume of size 120 × 187 × 51 by applying pyramid scheme and AOS algorithm, 20 equally spaced

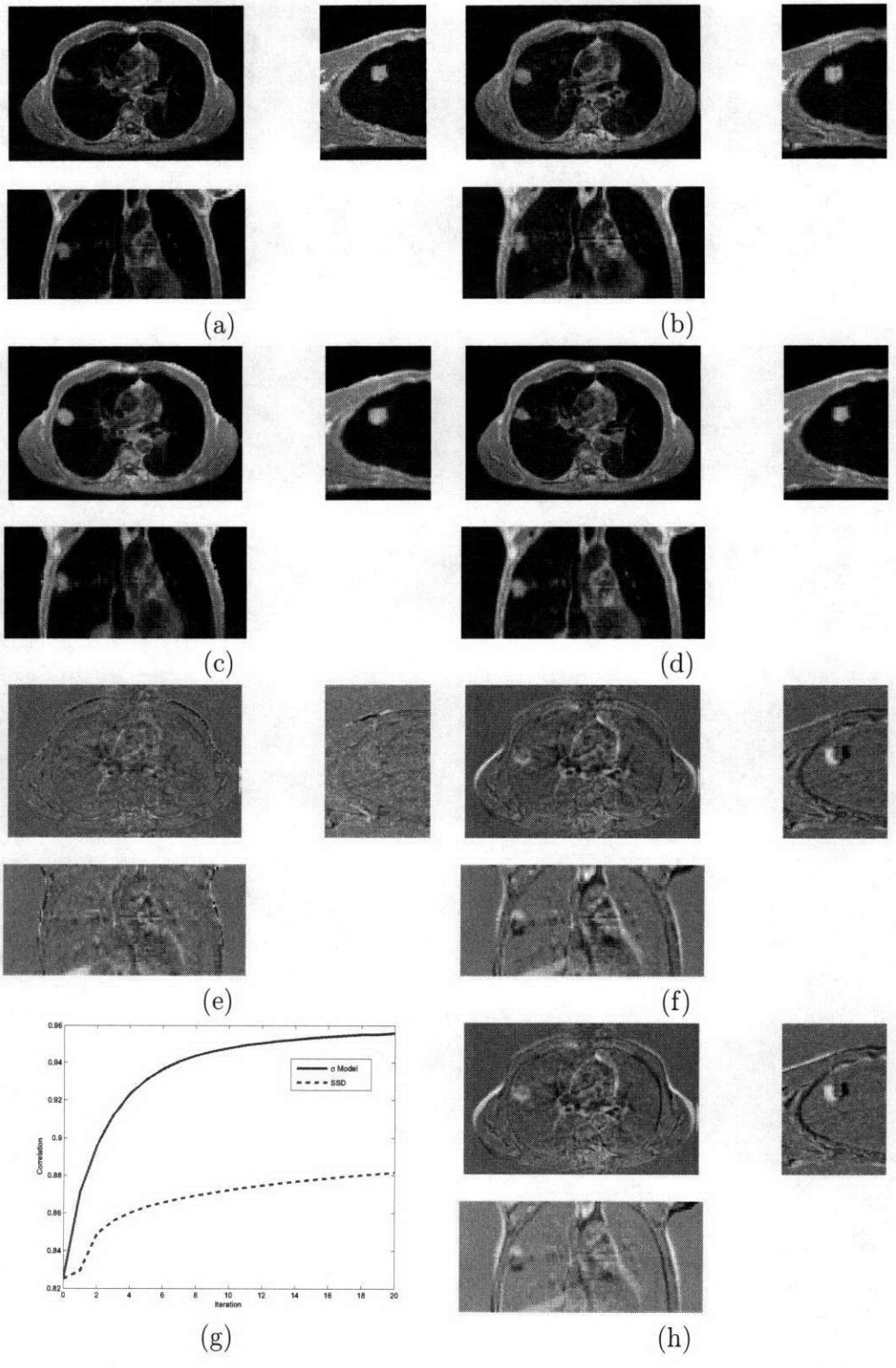


Figure 6-3. Registration for 3D MRI data in three views. (a,b)Source and target images respectively; (c,d) registration results by model (6–3) and model (6–1) respectively. (e,f) difference $T(X) - S(X + U(X))$ based on registration results by model (6–3) and model (6–1) respectively. (g) correlation; (h) original difference $T(X) - S(X)$

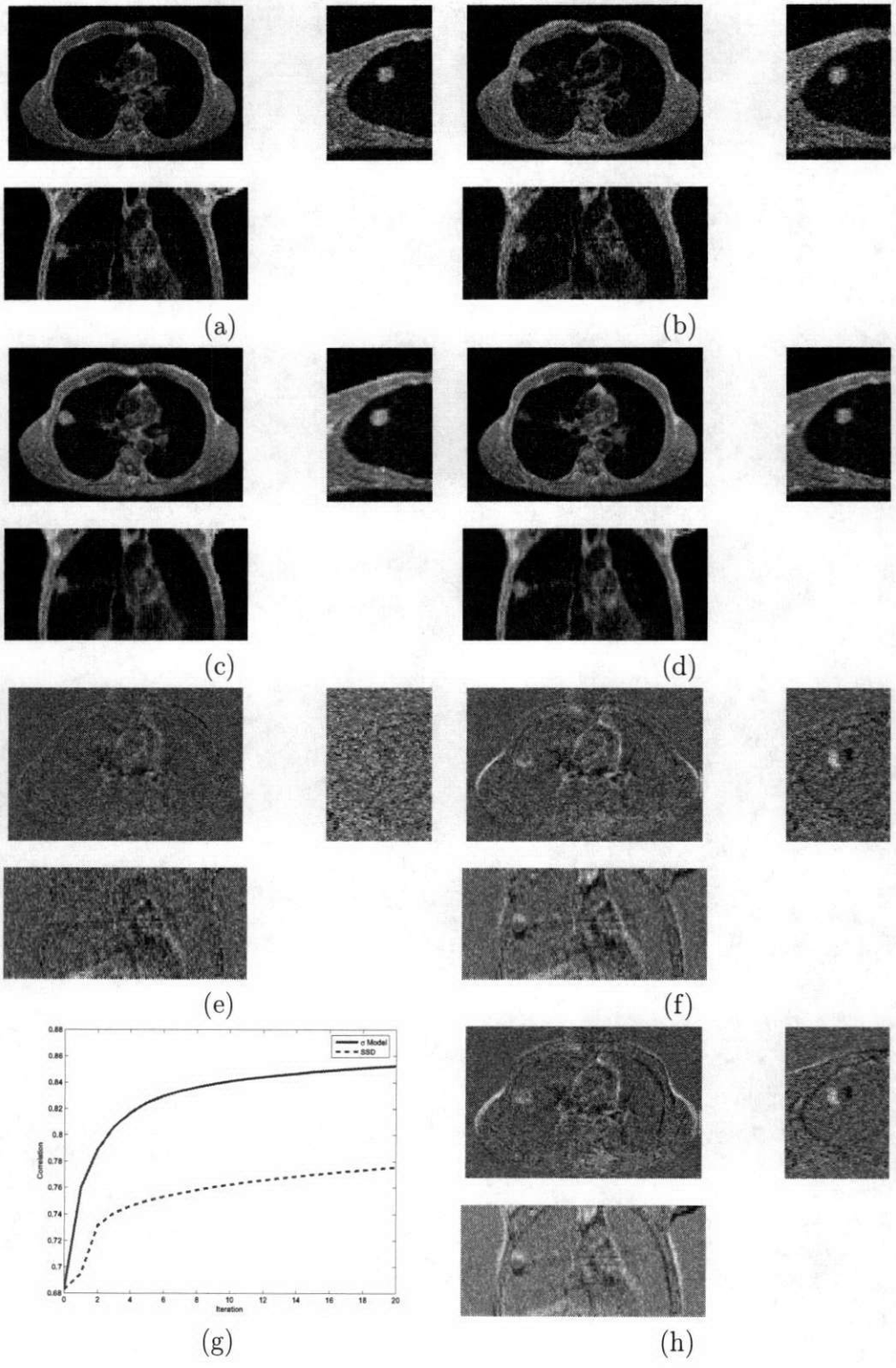


Figure 6-4. Registration of noised ($N(0, 0.005)$) 3D MRI data in three views. (a,b)Source and target images respectively; (c,d) registration results by model (6–3) and model (6–1) respectively. (e,f) difference $T(X) - S(X + U(X))$ based on registration results by model (6–3) and model (6–1) respectively. (g) correlation; (h) original difference $T(X) - S(X)$

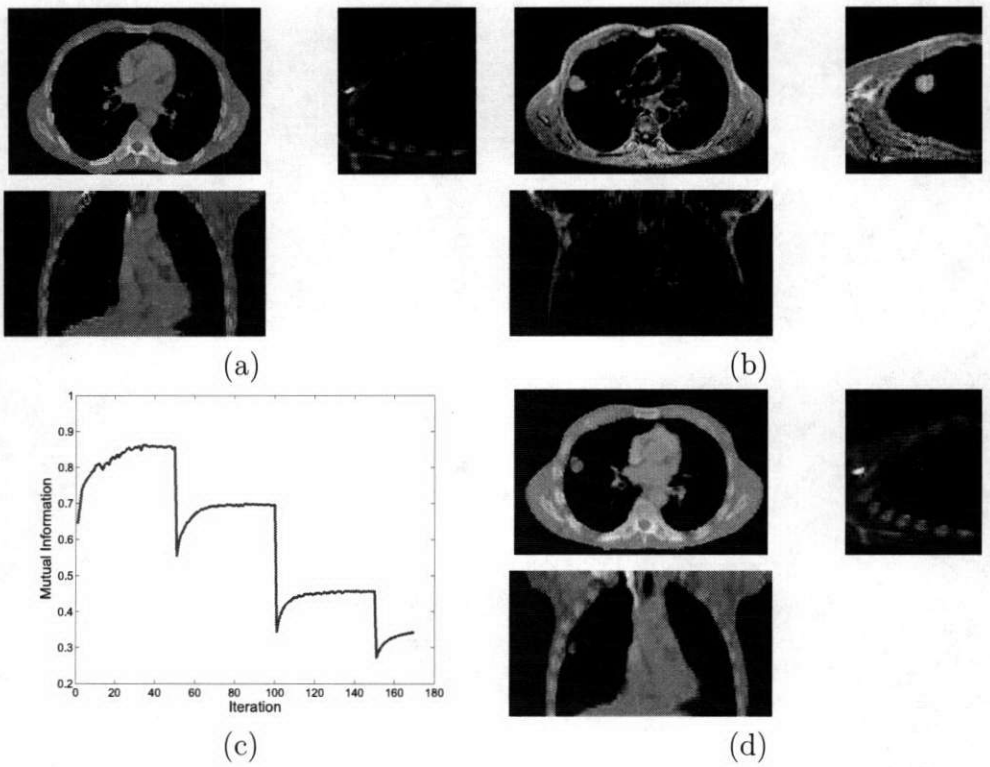


Figure 6-5. Registration of 3D CT volume to 3D MRI volume by model (6–7). (a,b)Source (CT) and target (MRI) images respectively; (c) Curve of MI during iterations; (d) Deformed result in three views.

bins are used during the iterations, and each iteration in the full resolution takes about 30 seconds on a laptop with Intel Core 2 Duo 1.66GHz CPU, 1GB RAM. Comparing with the source volume (a), one can see that the deformed volume in (d) is closer to the target volume (b), especially from the axial view of these three volumes. (c) shows that value of MI is increasing during iterations within each pyramid level.

6.5 Conclusion And Future Work

The proposed model (6–3) is featured at allowing the deformed template having a variance from the reference. This is especially good for aligning two images whose intensity are not exactly equal or linearly related, or corrupted by noise. Moreover, the weight between the smoothness and similarity measures is adjusted automatically in our model by the variance at each iteration. The weight on the goodness of matching increases when the standard variance decreases. This feature makes model (6–3) less sensitive than the model (6–1) to the choices of the parameter, this will be invested in the near future. All these features have contributed to the improvement of robustness and accuracy in registration greatly, as shown in experiments.

Due to clinical application requirement, efficient algorithms are needed to speed up the computation of MI in the multimodality case. Future work will focus on modeling and algorithms for robust and efficient registration.

CHAPTER 7
ACCURATE INVERSE CONSISTENT NON-RIGID IMAGE REGISTRATION AND ITS APPLICATION ON AUTOMATIC RE-CONTOURING

The main contribution of this chapter is on the improvement of inverse consistency, and its application to the radiation therapy, in particular, to get more accurate auto re-contouring. The basic idea is minimizing E_1 and E_2 in (1–14) coupled by the inverse consistent constraints defined in the next section. Applications of these inverse consistent deformations on auto re-contouring will be discussed in experimental results.

7.1 Proposed Method

In this section, we will first introduce a "natural" formulation of inverse consistent constraints, which can be used to correct the truncation errors in (1–17). Then we will combine this with the dissimilarity measures based on the likelihood of the residue image into our energy functionals to improve the accuracy, robustness and inverse consistent of the deformable image registration.

Let u and v be the forward and backward displacement fields related to deformation h and g by

$$h(\mathbf{x}) = \mathbf{x} + \mathbf{u}(\mathbf{x}), \quad g(\mathbf{x}) = \mathbf{x} + \mathbf{v}(\mathbf{x}). \tag{7–1}$$

Then the inverse consistent constraint $h(g(\mathbf{x})) = \mathbf{x}$ can be written in terms of u and v as

$$\mathbf{x} = h(g(\mathbf{x})) = g(\mathbf{x}) + u(g(\mathbf{x})) = \mathbf{x} + v(\mathbf{x}) + u(\mathbf{x} + v(\mathbf{x})). \tag{7–2}$$

Therefore,

$$v(\mathbf{x}) = -u(\mathbf{x} + v(\mathbf{x})) \tag{7–3}$$

Similarly, the constraint $g(h(\mathbf{x})) = \mathbf{x}$ can be represented by

$$u(\mathbf{x}) = -v(\mathbf{x} + u(\mathbf{x})) \tag{7–4}$$

In this work, we will use (7–3) and (7–4) as hard constraints in our proposed energy minimization method.

As in the chapter 6, to accommodate certain degree of variability in the image matching, we consider the residue between the deformed source image and target image $S(h(\mathbf{x})) - T(\mathbf{x})$ at each point as an independent random variable with Gaussian distribution of mean zero and a variance σ to be optimized. By the independency assumption, the joint pdf of all these random variables, which is the likelihood of the residual image given parameter σ, becomes

$$\begin{aligned} p(\{S(h(\mathbf{x})) - T(\mathbf{x}), \mathbf{x} \in \Omega\}|T(\mathbf{x}), \sigma) =& \Pi_{\mathbf{x}\in\Omega} p(S(h(\mathbf{x})) - T(\mathbf{x})|T(\mathbf{x}), \sigma) \\ =& \Pi_{\mathbf{x}\in\Omega} \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{|S(h(\mathbf{x}))-T(\mathbf{x})|^2}{2\sigma^2}}. \end{aligned} \tag{7–5}$$

Then the negative log-likelihood function is

$$\int_\Omega \frac{|S(h(\mathbf{x})) - T(\mathbf{x})|^2}{2\sigma^2} d\mathbf{x} + |\Omega| \ln \sqrt{2\pi}\sigma \tag{7–6}$$

Replace $M(S(h), T)$ in E_1 of (1–14) by (7–6), and replace $M(S, T(g))$ in E_2 in a similar manner, E_1 and E_2 can be rewritten in terms of u and v as:

$$\begin{aligned} E_1(u) &= \int_\Omega \frac{|S(\mathbf{x} + u(\mathbf{x})) - T(\mathbf{x})|^2}{2{\sigma_1}^2} d\mathbf{x} + |\Omega| \ln \sigma_1 + \lambda R(u(\mathbf{x})) \\ E_2(v) &= \int_\Omega \frac{|S(\mathbf{x}) - T(\mathbf{x} + v(\mathbf{x}))|^2}{2{\sigma_2}^2} d\mathbf{x} + |\Omega| \ln \sigma_2 + \lambda R(v(\mathbf{x})) \end{aligned} \tag{7–7}$$

After choosing $R(\bullet) = |\bigtriangledown \bullet|^2_{L^2(\Omega)}$ with boundary and initial conditions:

$$\begin{aligned} &\frac{\partial u}{\partial \vec{n}}(x,t) = 0, \ \frac{\partial v}{\partial \vec{n}}(x,t) = 0, \ \text{ on } \partial\Omega \times R^+ \\ &u(x,0) = 0, \ v(x,0) = 0, \ \text{ on } \Omega, \end{aligned} \tag{7–8}$$

ξ_1 and η_2 in (1–16) become

$$\begin{aligned} \eta_1(x) &= \frac{T(x) - S(x + u(x))}{\sigma_1^2} \nabla S(x + u(x)) + \lambda \triangle u(x) \\ \xi_2(x) &= \frac{S(x) - T(x + v(x))}{\sigma_2^2} \nabla T(x + v(x)) + \lambda \triangle v(x) \end{aligned} \tag{7–9}$$

Taking the first variation of the energy functional (7–7), we get

$$\sigma_1^2 = \frac{\int_\Omega |S(x+u(x)) - T(x)|^2 dx}{|\Omega|},\ \sigma_2^2 = \frac{\int_\Omega |S(x) - T(x+v(x))|^2 dx}{|\Omega|}. \tag{7–10}$$

By using this dissimilarity measure, the residual image no longer needs to be pointwisely close to zero to make the L^2 norm small. Instead, the new measure only forces the mean of the residue to be zero, and allows the residue having a variance to accommodate certain variability. This is especially good for aligning two images whose intensities are not exactly equal or linearly related, and makes the model more robust to noise and artifacts.

Moreover, the likelihood based approach is less sensitive to the choice of the parameter λ. In SSD models, λ is prefixed, so the balance of the dissimilarity measure and regularity measure does not change during iterations. This makes the selection of λ very difficult and affects registration result. In the proposed model the balancing factor of these two measures is, in fact, $\lambda\sigma^2$ rather than λ alone. Therefore, even λ is prefixed, the weight between these two measures varies at each iteration as the variance updates. As the iterations gradually approach to convergence stage, the residue magnitude becomes smaller, hence, the variance σ reduces, and consequently, the weight on smoothing deformation field versus matching images automatically decreases.

Combining these ideas, we propose a new model to improve the inverse consistency of (1–14) by using the hard constraints (7–3) and (7–4) to replace the penalty terms in (1–14), and to improve the efficiency of alignment by using the proposed similarity measure. More precisely, we propose to minimize a coupled minimization problem

$$\begin{cases} min_{u\in W^{1,2}(\Omega),\sigma_1}\ E_1(u,\sigma_1) \quad \text{and} \quad min_{v\in W^{1,2}(\Omega),\sigma_2} E_2(v,\sigma_2) \\ \qquad\qquad \text{where } E_1, E_2 \text{ are defined in (7–7)} \\ \qquad\qquad subject\ to\ u(x) + v(x+u(x)) = 0\ \forall x \in \Omega \\ \qquad\qquad v(x) + u(x+v(x)) = 0\ \forall x \in \Omega \end{cases} \tag{7–11}$$

Note if we only minimize $E_1(u)$, the solution u may not be invertible([103]), thus we may not be able to get its inverse through (7–3). By choosing η_1 and ξ_2 as in (7–9), the solutions of the constrained coupled minimization problem (7–11) are obtained via the following algorithm

Algorithm 1

$u(x) = 0; v(x) = 0; iter = 0;$

$corr = correlation(S, T)$

while $iter < maxstep$ and $corr < threshhold$ **do**

$iter++;$

$\mathbf{u_{new}} = \mathbf{u} + \mathbf{dt}\eta_1(\mathbf{u})$

$\mathbf{v} \leftarrow -\mathbf{u_{new}}(\mathbf{x} + \mathbf{v})$

$\mathbf{v_{new}} = \mathbf{v} + \mathbf{dt}\xi_2(\mathbf{v})$

$\mathbf{u} \leftarrow -\mathbf{v_{new}}(\mathbf{x} + \mathbf{u_{new}})$

$\mathbf{v} = \mathbf{v_{new}}$

$corr1 = correlation(S(x + u(x)), T(x))$

$corr2 = correlation(S(x), T(x + v(x)))$

$corr = min(corr1, corr2)$

end while

By applying the flow equations alternatively with the exact inverse consistent constraints in Algorithm 1, the forward and backward deformation forces are related via the constraints. In this manner, images to be registered are aligned well, meanwhile, the inverse inconsistency errors are controlled by hard constraints. The performance on image matching and improvement of accuracy on the inverse consistency will be shown in the next section.

To enhance the efficiency of the proposed algorithm, the Additive Operator Splitting scheme([89, 104]) is used to speed up our numerical computation. We present our results based on synthetic images, and 3D prostate MRI data.

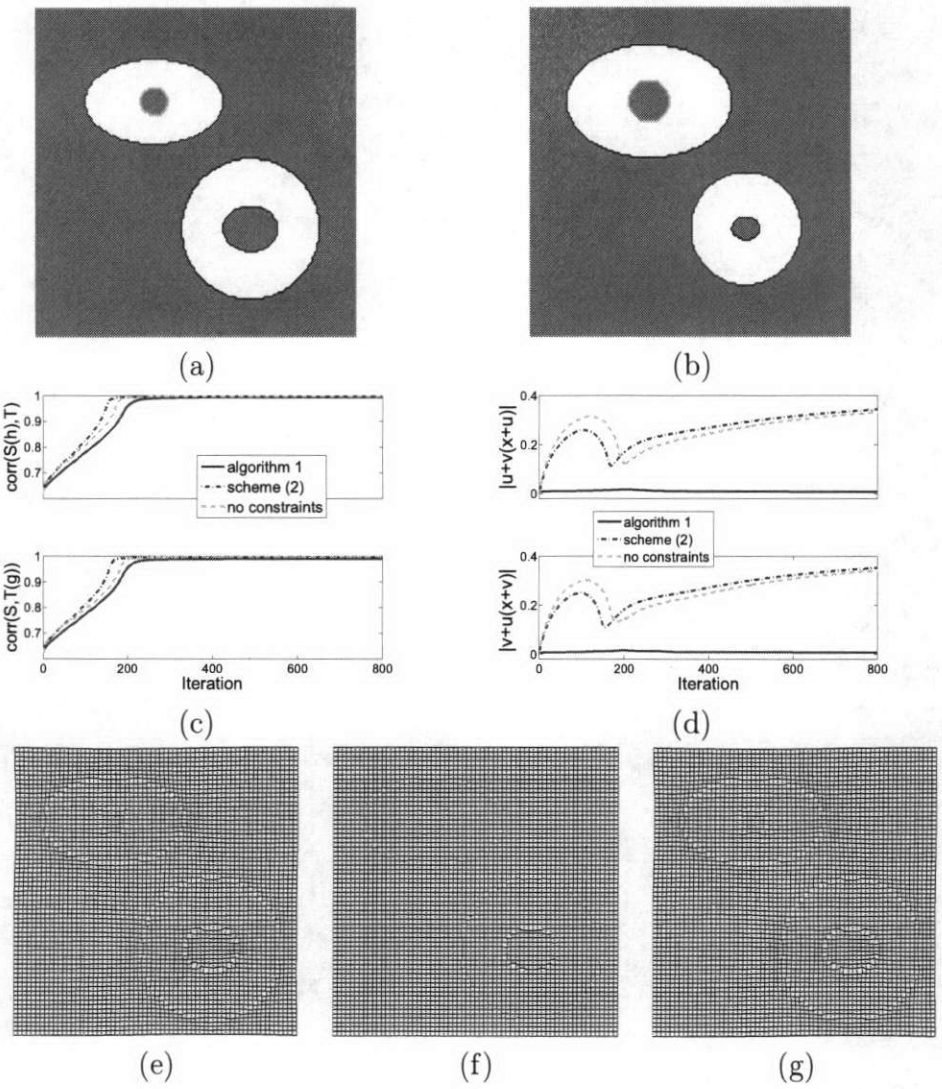


Figure 7-1. Experiment on 2D synthetic images. (a) Source with object contours superimposed, (b) target with contours superimposed, (c)correlation $Corr(S(h), T)$ and $Corr(S, T(g))$ during iterations, (d) mean of norms $\|u(x) + v(x + u(x))\|_{L^2(\Omega)}$ and $\|v(x) + u(x + v(x))\|_{L^2(\Omega)}$ during iterations, (e-g) grid representations of inverse inconsistent error field $u(x) + v(x + u(x))$ by Scheme (1–15), Algorithm 1 with and without constraints respectively.

7.2 Experiment Results

In this section, we show our experimental results on 2D synthetic images, 3D prostate MRI data, and 3D CT data, which indicate the improvement of the proposed algorithm in accuracy of inverse consistency and accuracy of auto re-contouring.

Based on (7–3) and (7–4), if h and g are inverse to each other, then both of the inverse inconsistent error fields $u(\mathbf{x}) + v(\mathbf{x} + u(\mathbf{x}))$ and $v(\mathbf{x}) + u(\mathbf{x} + v(\mathbf{x}))$ should be 0. Thus, their components and norms should also be 0. This property will be used to compare the inverse inconsistency errors between three methods: Scheme (1–15), Algorithm 1 and Model (7–11) without constraints.

The first experiment is aimed to exam and compare the inverse consistency of these three methods on synthetic data. Fig 7-1(a) and (b) present the source image S and target image T, respectively, with the boundaries of the objects superimposed. Objects have intensity 1.0 in both images, and their background/holes have intensity 0. The three methods are applied separately with the same parameters $dt = .05, \lambda = 5.0$ for 800 iterations, and the corresponding results are shown in Fig 7-1(c-g). From Fig 7-1(c), one can see that correlation by all methods converges to about 1.0 similarly. However in Fig 7-1(d), the means of norms for both inverse inconsistent error fields $u(\mathbf{x}) + v(\mathbf{x} + u(\mathbf{x}))$ and $v(\mathbf{x}) + u(\mathbf{x} + v(\mathbf{x}))$ by both Scheme (1–15) and non-constrained Model (7–11) are increased to about 0.4 pixels in average, i.e., the mean value of their inverse inconsistency errors are about 0.4 pixels, while those from Algorithms 1 are maintained in a negligible low level (about 0.01 pixel). We also compare the error field $u(\mathbf{x}) + v(\mathbf{x} + u(\mathbf{x}))$ in Fig 7-1(e,f,g) by applying results from Scheme (1–15), Algorithm 1 and Model (7–11) without constraints respectively on a regular grid mesh. Fig 7-1(f) shows that the error by Algorithm 1 is 0 almost everywhere(almost no displacement in the regular grid mesh), while (e) and (g)

Table 7-1. Inverse inconsistency error comparison results for synthetic images: the components and norms of inverse inconsistency error fields $u(\mathbf{x}) + v(\mathbf{x} + u(\mathbf{x}))$ and $v(\mathbf{x}) + u(\mathbf{x} + v(\mathbf{x}))$. X_{max}, X_{mean} denote the maximum and mean values of the first component of the error fields respectively, Y_{max}, Y_{mean} denote that for the second component. $\| \bullet \|$ denotes norms of the error fields at each pixel. Ω is the image domain

inconsistency error $u(\mathbf{x}) + v(\mathbf{x} + u(\mathbf{x}))$ on Ω						
Method	X_{max}	X_{mean}	Y_{max}	Y_{mean}	$\| \bullet \|_{max}$	$\| \bullet \|_{mean}$
Algorithm 1	0.7396	0.0051	0.7876	0.0059	0.7879	0.0085
Scheme (1–15)	0.9915	0.2390	1.1055	0.2320	1.2272	0.3526
No constraints	0.9916	0.2300	1.0979	0.2255	1.2202	0.3414
inconsistency error $v(\mathbf{x}) + u(\mathbf{x} + v(\mathbf{x}))$ on Ω						
Algorithm 1	0.4870	0.0047	0.5814	0.0053	0.7287	0.0076
Scheme (1–15)	0.8705	0.2290	1.2224	0.2240	1.2304	0.3398
No Constraints	0.8303	0.2199	1.2670	0.2174	1.2729	0.3284
inconsistency error $u(\mathbf{x}) + v(\mathbf{x} + u(\mathbf{x}))$ on contours in S						
Algorithm 1	0.2810	0.0431	0.2769	0.0502	0.3742	0.0719
Scheme (1–15)	0.7237	0.1445	0.8047	0.1878	0.8073	0.2577
No Constraints	0.7403	0.1461	0.8374	0.1908	0.8410	0.2612
inconsistency error $v(\mathbf{x}) + u(\mathbf{x} + v(\mathbf{x}))$ on contours in T						
Algorithm 1	0.3833	0.0412	0.4058	0.0484	0.4059	0.0697
Scheme (1–15)	0.5937	0.1390	0.9308	0.1661	0.9715	0.2361
No Constraints	0.6180	0.1410	0.9735	0.1695	1.0171	0.2402

Table 7-2. Inverse inconsistency error comparisons for the $1st$ and $21st$ phases of a 3D prostate MRI data on regions Ω and all the contours in the $1st$ phase S.

Method	X_{max}	X_{mean}	Y_{max}	Y_{mean}	$\|\bullet\|_{max}$	$\|\bullet\|_{mean}$
inconsistency error field $u(\mathbf{x}) + v(\mathbf{x} + u(\mathbf{x}))$ on Ω						
Algorithm 1	0.5155	0.0105	0.4010	0.0082	0.5549	0.0146
Scheme (1–15)	0.4936	0.0457	0.6750	0.0793	0.7154	0.1020
inconsistency error field $v(\mathbf{x}) + u(\mathbf{x} + v(\mathbf{x}))$ on Ω						
Algorithm 1	0.5937	0.0121	0.4766	0.0087	0.6230	0.0164
Scheme (1–15)	1.6315	0.0583	1.5860	0.0951	1.9917	0.1231
inconsistency error $u(\mathbf{x}) + v(\mathbf{x} + u(\mathbf{x}))$ on contours in S						
Algorithm 1	1.6165	0.0524	1.9630	0.0459	2.3577	0.0770
Scheme (1–15)	4.7525	0.1036	6.2510	0.1292	6.3180	0.1843

show errors from other two methods are larger, especially in the region corresponding to the boundaries of two holes in S and T.

To quantitatively validate the improved inverse consistency by the proposed algorithm, we compare the maximum and mean values of components and norms of $u(\mathbf{x}) + v(\mathbf{x} + u(\mathbf{x}))$ and $v(\mathbf{x}) + u(\mathbf{x} + v(\mathbf{x}))$ in Table 7-1, where X_{max}, X_{mean} denote the maximum and mean values of the first component of the error fields respectively, and Y_{max}, Y_{mean} are those of the second components. The quantitative comparisons are performed in two regions: one is in the image domain Ω, and the other is on the contours which are the boundaries of the objects in the images shown in Fig 7-1(a,b) respectively. Table 1 shows the proposed algorithm yields much smaller errors in all aspects. Particularly, its mean error is about one fortieth of those from both scheme (1–15) and non-constrained model (7–11) in Ω, and about one thirtieth of theirs at the contour regions.

The second experiment is to validate the improvement in accuracy of the proposed algorithm on 3D prostate data, which consists of 100 phases of 2D images focusing on prostate area, where ROIs have large internal motions. The source volume S is the first phase, and the boundaries of ROIs in S are delineated by contours and superimposed in Fig 7-2(a), and the other 99 phases are targets, and methods Scheme (1–15) and Algorithm 1 are applied to find deformations between the $1st$ phase and each of the

other 99 phases. The auto re-contouring, that register the contours in S into the other 99 phases, is achieved by applying the deformations on these contours. For demonstration purpose, one of the target image T, the $21st$ phase, is shown in Fig 7-2(b), and the auto re-contouring result by Algorithm 1 is superimposed on it.

Comparison on the convergence and inverse inconsistency by the first two methods are shown in Fig 7-2(c,d). Fig 7-2(c) compares CC between deformed images and target images by these two models. Both Scheme (1–15) and Algorithm 1 improve the initial CC between S and each of other 99 phases to a similar level. However, from Fig 7-2(d) we can observe that the norm of inverse inconsistency error by scheme (1–15) is much higher in average than that of Algorithm 1.

The quantitative comparison on inverse inconsistency errors between the $1st$ phase and the 21st phase for this experiment is listed in Table 7-2 for demonstration. Beside comparing the error fields on Ω, we also evaluate error $v(x)+u(x+v)$ at points of all given contours on S. By comparing the corresponding components and norms of the inverse inconsistency error fields in Table 2, we find that errors generated by proposed algorithm is much lower than that by scheme (1–15), this indicates that the point correspondence and automatic re-contouring results are more accurate.

The third experiment is to validate the improved accuracy of proposed algorithms on 3D CT data. The source volume S, which is treated as planning data in a radiation therapy, is shown in Fig 7-3(a) in three views (the axial, sagittal and coronal views). The boundaries of organs in S are drawn slice by slice in axial view as shown in Fig 7-3(a). The target volume T is shown in Fig 7-3(b). Both S and T are of dimension $256\times256\times64$. We apply the three algorithms on this data as in the first experiment. A multi-resolution approach is used in our computation. We start from resolution $64\times64\times16$ with parameters $\lambda = 10$, $dt = 0.4$ for 40 iterations, then increase the resolution to $128 \times 128 \times 32$ with $dt = .2$ for 20 iterations, and finally to the original resolution with $dt = .1$ for 10 steps.

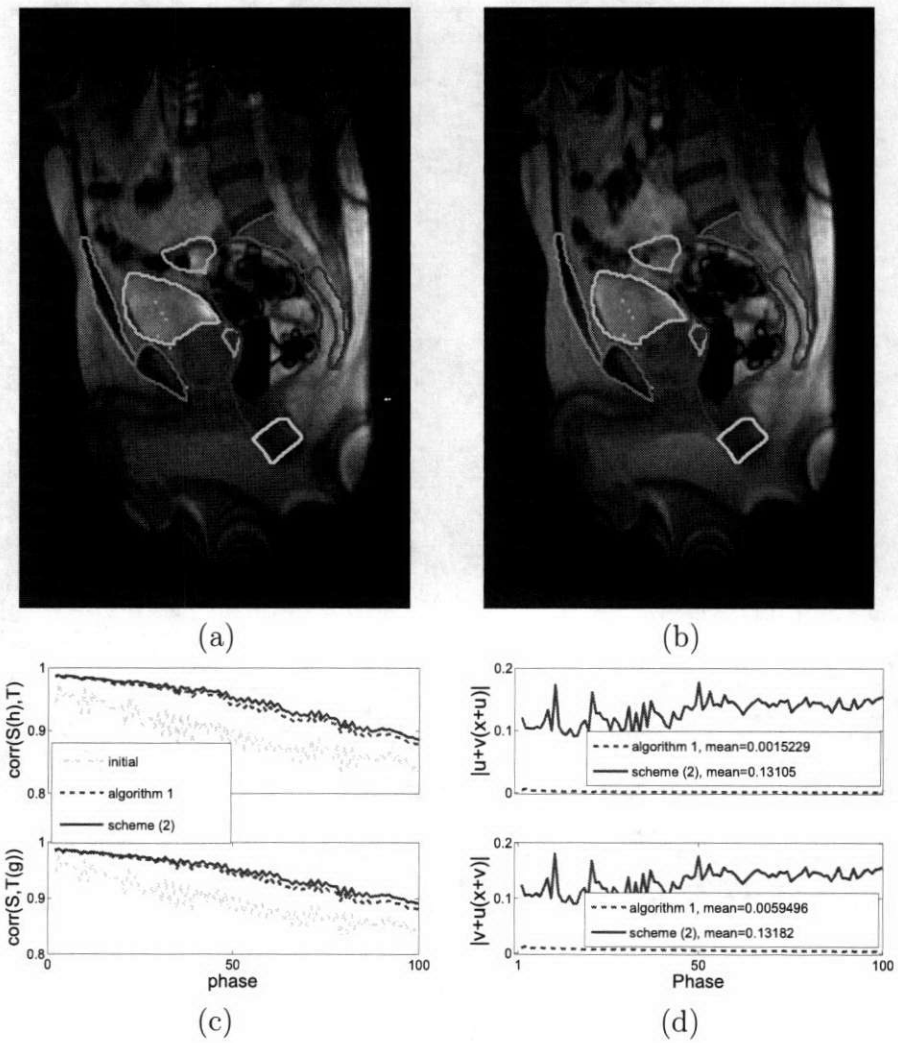


Figure 7-2. Experiment on 3D prostate MRI data with 100 2D phases. (a) Source image(the $1st$ phase) with contours of ROIs superimposed, (b) the $21st$ phase with automatic recontouring results by Algorithm 1 superimposed, (c) plots of correlation $Corr(S(h), T)$, and $Corr(S, T(g))$ between the $1st$ phase and each of other 99 phases, (d) plots of norms $||u(x) + v(x + u(x))||$ and $||v(x) + u(x + v(x))||$ for deformations between the $1st$ phase and each of other 99 phases, (c,d) are based on parameters $\lambda = 20, dt = .1$ for 200 iterations

The tri-linearly interpolated u and v from the previous lower resolution are used as the initials for the next finer resolution.

Comparison on the convergence and inverse inconsistency by the three methods are shown in Fig 7-3(c,d). Fig 7-3(c) compares cross correlation (CC) of these three models. Scheme (1–15) converges slightly faster than the other two, but finally the CC tends to the same level. However, from Fig 7-3(d) we can observe that the norm of inverse inconsistency error by scheme (1–15) is much higher in average than that of Algorithm 1. This again shows the improvement of inverse consistency of the proposed methods.

The quantitative comparison on inverse inconsistency errors for the third experiment is listed in Table 7-3. By comparing the corresponding components and norms of the inverse inconsistency error fields in Table 7-3, we find that errors generated by proposed algorithms is about one fifth of that by scheme (1–15).

7.3 Conclusion

In this work, we proposed a coupled energy minimization method with inverse consistent constraints for deformable image registration. The proposed model controls the inverse inconsistency errors in a negligibly low level, therefore, it provides a better correspondence for the ROIs in source and target images. This makes the auto re-contouring results with data involved in the course of radiation therapy much more accurate.

The dissimilarity measure used in this work is the negative log-likelihood of the residual image between the deformed source and target. This dissimilarity measure is able to accommodate certain variability in the matching. Hence, the model is more robust to noise than SSD, moreover, it is less sensitive to the choice of the parameter that balances the smoothness of the deformation field and goodness of matching.

Table 7-3. Inverse consistency error comparisons for 3D CT data. Z denotes the third component of the inverse inconsistency error fields, based on regions: Ω, and all contours associated with the source volume S.

inconsistency error field $u(\mathbf{x}) + v(\mathbf{x} + u(\mathbf{x}))$ on Ω								
Method	X_{max}	X_{mean}	Y_{max}	Y_{mean}	Z_{max}	Z_{mean}	$\|\bullet\|_{max}$	$\|\bullet\|_{mean}$
Algorithm 1	8.9336	0.0151	3.0239	0.0154	1.4882	0.0084	9.0485	0.0265
Scheme (1–15)	6.6532	0.0740	3.6780	0.0765	2.2982	0.0734	7.1051	0.1523
inconsistency error field $v(\mathbf{x}) + u(\mathbf{x} + v(\mathbf{x}))$ on Ω								
Algorithm 1	5.3978	0.0135	2.8936	0.0148	2.9918	0.0079	5.4209	0.0247
Scheme (1–15)	6.3512	0.0724	2.7737	0.0760	4.7697	0.0722	7.9680	0.1499
inconsistency error field $v(\mathbf{x}) + u(\mathbf{x} + v(\mathbf{x}))$ on all contours in S								
Algorithm 1	1.9935	0.0210	1.0991	0.0169	0.8705	0.0138	2.5338	0.0320
Scheme (1–15)	2.5347	0.0853	1.3701	0.0770	0.8306	0.0984	3.0055	0.1491

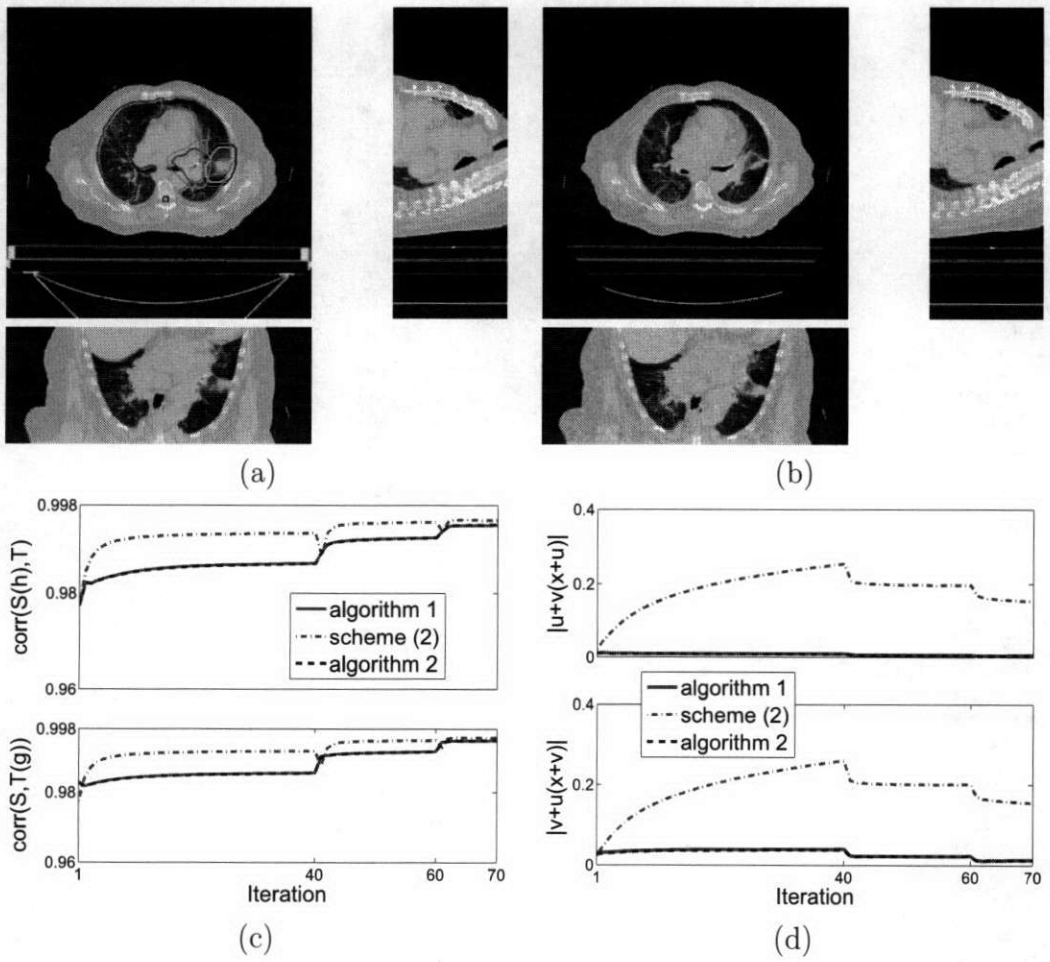


Figure 7-3. Experiment on 3D CT volume. (a) Source volume with contours supperimposed, (b) target volume, (c) correlations $corr(S(h), T)$, and $corr(S, T(g))$ during iterations, (d) mean of $||u(x) + v(x + u(x))||$ and $||v(x) + u(x + v(x))||$ during iterations.

Acknowledgments

I would like to thank Dr. Jim Dempsey, Dr. Jonathan Li and Dr. Anneyuko I. Saito from Department of Radiation Oncology at University of Florida and Viewray Inc. for providing the CT Lung and prostate data.

CHAPTER 8
FUTURE WORK

In this chapter, I will briefly discuss my future research directions related with my work presented in previous chapters.

8.1 Diffusion Weighted Magnetic Resonance Image Analysis

In last few years, an active research area in diffusion weighted image is to estimate the diffusion Orientation Distribution Function (ODF) (e.g. [105–107]). In [105],Tuch proposed Q-Ball Imaging, where the ODF is estimated directly from the raw HARD MRI measurements on a single sphere by the Funk-Radon Transform (FRT), which was later solved analytically and efficiently by Descoteaux et. al. in [106] by expressing the HARD MRI signal as a spherical harmonic series up to some order and by using the Funk-Hecke theorem [108]. McGraw et. al. [107] proposed a von Mises-Fisher mixture model of the ODF. Jian et. al. in [109] reconstructed multi-fiber by using Mixture of Wisharts and sparse deconvolution. Recent work regarding fiber trace reconstruction based on ODF can be found in [110, 111].

8.2 Deformable Image Registration

Image registration has been extensively studied in last few decades, but due to massive data involved in most medical applications, it's still very challenging to make these processes efficient. While real time performance is an essential requirement for some critical clinical applications like image guided radiation therapy. My future work will also focus on efficient algorithm developing on image registration including both intensity based and feature based image registration techniques regarding both same-modality and multi-modality data(e.g. [74, 112–114]), via novel models or even through different hardware architectures (e.g. [115, 116]).

Still another direction of my future work is on diffusion weighted MRI registration (e.g. [117, 118]), which can be done by choosing appropriate dissimilarity measures (e.g. [119, 120]) and translations.

APPENDIX A
SSD MODEL: EXISTENCE OF MINIMIZER

$$\min_{u\in W^{1,2}(\Omega)} E(u) := \frac{1}{2}\int_{\Omega}\{\lambda|\nabla u|^2 + |S(x+u(x)) - T(x)|^2\}dx, \qquad \text{(A–1)}$$

Theorem: Assume Ω is open and bounded in R^n, $S(x), T(x)$ are bounded in R^n, $|u|_{L^\infty(\Omega)} \leq C(|\Omega|)$ for some constant C depending only on size of Ω, then (A–1) has at least one solution.

Proof: It's obvious that $\forall u \in W^{1,2}(\Omega)$, $E(u) \geq 0$, bounded below, so there exists a minimizing sequence $\{u_k\} \subset W^{1,2}(\Omega)$ s.t.

$$\lim_{k\to\infty} E(u_k) \;=\; inf_{u\in W^{1,2}(\Omega)} E(u) := m.$$

Hence there exists a $K \in N$ such that $\forall k > K$, $E(u_k) \leq m+1$. so $\{E(u_k)\}$ is bounded. By the fact that $E(u)$ is a sum of positive terms, therefore $\{\|Du_k\|_{L^2(\Omega)}\}$ is bounded. $|u|_{L^\infty(\Omega)} \leq C$ implies $\{\|u_k\|_{L^2(\Omega)}\}$ bounded, so $\{u_k\}$ is bounded in $W^{1,2}(\Omega)$.

$W^{1,2}(\Omega)$ is reflexive, so there exists a subsequence $\{u_{k_j}\} \subset \{u_k\}$ and $u \in W^{1,2}(\Omega)$, such that

$$u_{k_j} \rightharpoonup u \; in \; W^{1,2}(\Omega).$$

Clearly, $E(u)$ is bounded below, and convex w.r.t ∇u, so $E(u)$ is weak lower semi-continuous on $W^{1,2}(\Omega)$. i.e:

$$E(u) \leq \liminf_{j\to\infty} E(u_{k_j}) = \liminf_{k\to\infty} E(u_k) = m$$

While $u \in W^{1,2}(\Omega)$,

$$E(u) \geq \liminf_{k\to\infty} E(u_k) = m,$$

so

$$E(u) = \liminf_{k\to\infty} E(u_k) = m,$$

then u is a minimizer of (A–1)¶

APPENDIX B
σ MODEL: EXISTENCE OF MINIMIZER

$$\begin{aligned} E(u,\sigma) &:= \int_\Omega \frac{|S(x+u(x))-T(x)|^2}{2\sigma^2}dx + |\Omega|\ln\sigma + \frac{\lambda}{2}\int_\Omega |\bigtriangledown u(x))|^2 dx \\ &= \int_\Omega \frac{|S(x+u(x))-T(x)|^2}{2\sigma^2} + \ln\sigma + \frac{\lambda}{2}|\bigtriangledown u(x))|^2 dx \end{aligned} \tag{B–1}$$

Theorem: Assume Ω is an open and bounded subset in R^n, $S(x), T(x)$ are continuous and bounded in R^n, $\|u\|_{L^\infty(\Omega)} \le C(|\Omega|)$. If $\exists \varepsilon > 0$ s.t.

$$\int_\Omega |S(x+u(x))-T(x)|^2 dx \ge \varepsilon,\ \forall\, u \in W^{1,2}(\Omega),$$

then (B–1) has at least one minimizer $(u,\sigma) \in W^{1,2}(\Omega)\times(0,\infty)$.

Proof:

1. let's prove $E(u,\sigma)$ is bounded below.

By assumption, there $\exists \varepsilon > 0$ such that $\int_\Omega |S(x+u(x))-T(x)|^2 dx \ge \varepsilon,\ \ \forall\, u \in W^{1,2}(\Omega)$, thus

$$\frac{1}{\sigma} e^{-\frac{\int_\Omega |S(x+u(x))-T(x)|^2 dx}{2\sigma^2}} \le \frac{1}{\sigma} e^{-\frac{\varepsilon}{2\sigma^2}}.$$

By the fact that

$$\frac{1}{\sigma} e^{-\frac{\varepsilon}{2\sigma^2}} \le \frac{1}{\sqrt{\varepsilon}} e^{-\frac{1}{2}},\ \text{where "=" holds as } \sigma^2 = \varepsilon.$$

We have

$$\begin{aligned} \int_\Omega \frac{|S(x+u(x))-T(x)|^2}{2\sigma^2} + \ln\sigma dx &= -ln\frac{1}{\sigma} e^{-\frac{\int_\Omega |S(x+u(x))-T(x)|^2 dx}{2\sigma^2}} \\ &\ge -ln\frac{1}{\sigma} e^{-\frac{\varepsilon}{2\sigma^2}} \\ &\ge -\frac{1}{2}(1+\ln\varepsilon) \end{aligned} \tag{B–2}$$

so $E(u,\sigma)$ is bounded from below.

2. So there exists a minimizing sequence $\{u_k,\sigma_k\} \subset W^{1,2}(\Omega)\times(0,\infty)$ s.t.

$$\lim_{k\to\infty} E(u_k,\sigma_k) \ = \ inf_{u\in W^{1,2}(\Omega),\sigma>0} E(u,\sigma) := m. \tag{B–3}$$

Rewrite

$$m = m_1 + m_2 + m_3$$

where m_1, m_2, m_3 are corresponding values for $\int_\Omega \frac{|S(x+u(x))-T(x)|^2}{2\sigma^2}dx$, $|\Omega|\ln\sigma$ and $\frac{\lambda}{2}\int_\Omega |\bigtriangledown u(x))|^2 dx$ respectively at $arg\ inf_{u\in W^{1,2}(\Omega),\sigma>0} E(u,\sigma)$.

By (B-3), there exists $K_0 \in N$ such that $\forall k \geq K_0$, we have

$$m \leq E(u_k, \sigma_k) \leq m + 1.$$

Since

$$0 \leq \lim_{\sigma\to 0^+,\infty} \frac{1}{\sigma} e^{-\frac{\int_\Omega |S(x+u(x))-T(x)|^2 dx}{2\sigma^2}} \leq \lim_{\sigma\to 0^+,\infty} \frac{1}{\sigma} e^{-\frac{M}{2\sigma^2}} = 0,$$

We have

$$\lim_{\sigma\to 0^+,\infty} \int_\Omega \frac{|S(x+u(x))-T(x)|^2}{2\sigma^2} + \ln\sigma dx = +\infty \tag{B-4}$$

Hence for the given minimizing sequence $\{(u_k, \sigma_k)\} \subset W^{1,2}(\Omega) \times (0,\infty)$, $\sigma_k \nrightarrow 0$ and $\sigma_k \nrightarrow +\infty$. i.e., σ_k is bounded below and above, there exist $\delta \in (0,\infty)$ such that $\sigma_k \in [\delta, \frac{1}{\delta}]$ for k large enough $(> K_1)$.

3. Set $K_2 = max(K_0, K_1)$, then $\forall k \geq K_2$, σ_k is bounded, so is $ln\sigma_k$, while

$$\int_\Omega \frac{|S(x+u_k(x))-T(x)|^2}{2{\sigma_k}^2} dx \geq 0, \tag{B-5}$$

we have

$$\{\bigtriangledown u_k\} \subset L^2(\Omega) \text{ is bounded.}$$

By the assumption that $|u_k|_{L^\infty(\Omega)} \leq C(|\Omega|)$, we have $\{u_k\} \subset L^2(\Omega)$ is bounded , so $\{u_k\}$ is bounded in $W^{1,2}(\Omega)$. Hence, there exists a subsequence $\{u_k\} \subset W^{1,2}(\Omega)$ and $u \in W^{1,2}(\Omega)$ such that

$$u_k \rightharpoonup u \text{ weakly in } W^{1,2}(\Omega).$$

Since σ_k is bounded, there also exists a subsequence $\{\sigma_k\}$ and σ such $\sigma_k \to \sigma$.

We are to show

$$E(u,\sigma) \leq m.$$

i.e.

$$\lim_{k\to\infty}\int_\Omega \frac{|S(x+u_k(x))-T(x)|^2}{2{\sigma_k}^2}dx \le m_1 \tag{B-6}$$

$$\lim_{k\to\infty} |\Omega| \ln\sigma_k \le m_2 \tag{B-7}$$

$$\lim_{k\to\infty} \frac{\lambda}{2}\int_\Omega |\bigtriangledown u_k(x))|^2 dx \le m_3 \tag{B-8}$$

Clearly, since $ln(x)$ is continues, $\ln\sigma_k \to \ln\sigma$, (B–7) is proved.

$\int_\Omega |\bigtriangledown u(x))|^2 dx$ is convex w.r.t. $\bigtriangledown u$, so it is weak lower semi-continuous in $W^{1,2}(\Omega)$, i.e. (B–8) is proved.

The only thing left is to prove (B–6).

4. From compactness theorem, there exists a subsequence $\{u_{k_j}, \sigma_{k_j}\}$(we still denote it as $\{u_k, \sigma_k\}$) such that

$$u_k \to u \text{ a.e. in } \Omega \tag{B-9}$$

By Dominated Convergence Theorem,

$$\int_\Omega |S(x+u_k(x))-T(x)|^2 dx \to \int_\Omega |S(x+u(x))-T(x)|^2 dx \tag{B-10}$$

Hence,

$$
\begin{aligned}
&|\int_\Omega \frac{|S(x+u_k(x))-T(x)|^2}{2{\sigma_k}^2}dx-\int_\Omega \frac{|S(x+u(x))-T(x)|^2}{2\sigma^2}dx| \\
\leq&|\int_\Omega \frac{|S(x+u_k(x))-T(x)|^2}{2{\sigma_k}^2}dx-\int_\Omega \frac{|S(x+u(x))-T(x)|^2}{2{\sigma_k}^2}dx| \\
&+|\int_\Omega \frac{|S(x+u(x))-T(x)|^2}{2{\sigma_k}^2}dx-\int_\Omega \frac{|S(x+u(x))-T(x)|^2}{2\sigma^2}dx| \\
=&\frac{1}{2{\sigma_k}^2}|\int_\Omega |S(x+u_k(x))-T(x)|^2dx-\int_\Omega |S(x+u(x))-T(x)|^2dx| \\
&+\int_\Omega |S(x+u(x))-T(x)|^2dx|\frac{1}{2{\sigma_k}^2}-\frac{1}{2\sigma^2}| \\
\leq&\frac{1}{2\delta^2}|\int_\Omega |S(x+u_k(x))-T(x)|^2dx-\int_\Omega |S(x+u(x))-T(x)|^2dx| \\
&+\int_\Omega |S(x+u(x))-T(x)|^2dx|\frac{1}{2{\sigma_k}^2}-\frac{1}{2\sigma^2}| \\
\to&0.
\end{aligned}
\tag{B-11}
$$

Thus (B-6) is proved. Thus (u,σ) is a minimizer of (B-1).

APPENDIX C
INVERSE CONSISTENT MODEL: EXISTENCE OF MINIMIZER

$$E_1(u,\sigma_1) = \int_\Omega \frac{|S(\mathbf{x}+u(\mathbf{x})) - T(\mathbf{x})|^2}{2{\sigma_1}^2} d\mathbf{x} + |\Omega| \ln \sigma_1 + \frac{\lambda}{2}\int_\Omega |\bigtriangledown u(\mathbf{x}))|^2 d\mathbf{x}$$
$$E_2(v,\sigma_2) = \int_\Omega \frac{|S(\mathbf{x}) - T(\mathbf{x}+v(\mathbf{x}))|^2}{2{\sigma_2}^2} d\mathbf{x} + |\Omega| \ln \sigma_2 + \frac{\lambda}{2}\int_\Omega |\bigtriangledown u(\mathbf{x}))|^2 d\mathbf{x} \quad \text{(C–1)}$$

$$\begin{cases} min_{u\in W^{1,2}(\Omega),\sigma_1>0}\, E_1(u,\sigma_1) \quad \text{and} \quad min_{v\in W^{1,2}(\Omega),\sigma_2>0} E_2(v,\sigma_2) \\ \qquad subject\ to\ u(x) + v(x+u(x)) = 0\ a.e\ in\ \Omega \\ \qquad\qquad v(x) + u(x+v(x)) = 0\ a.e\ in\ \Omega \end{cases} \quad \text{(C–2)}$$

Theorem: Assume Ω is open and bounded in R^n, $S(x), T(x)$ are bounded in R^n, $|u|_{L^\infty(\Omega)}, |v|_{L^\infty(\Omega)} \le C$ for some constant C depending on Ω. If there $\exists \varepsilon > 0$ such that

$$\int_\Omega |S(x+u(x)) - T(x)|^2 dx \ge \varepsilon,\ \forall\ u \in W^{1,2}(\Omega)$$

and

$$\int_\Omega |S(x) - T(x+v(x))|^2 dx \ge \varepsilon,\ \forall\ v \in W^{1,2}(\Omega),$$

then (C–2) has at least one solution such that u, v in $\mathcal{A} := \{u, v \in W^{1,2}(\Omega) | u(x) = -v(x+u(x)\ a.e., v(x) = -u(x+v(x)\ a.e)\}$.

Proof:

First, $\mathcal{A}$ is not empty, since $(u \equiv 0, v \equiv 0) \in \mathcal{A}$

As in Appendix B, we have a minimizing sequence $\{(u_k, \sigma_{1k}, v_k, \sigma_{2k})\}_{k=1}^\infty$ with $\{(u_k, v_k)\}_{k=1}^\infty \in \mathcal{A}$ such that

$$E_1(u_k, \sigma_{1k}) \to m_1 = \inf_{u\in\mathcal{A},\sigma_1>0} E_1(u,\sigma_1) \quad \text{(C–3)}$$

$$u_k \rightharpoonup u \text{ weakly in } W^{1,2}(\Omega) \quad \text{(C–4)}$$

$$\sigma_{1k} \to \sigma_1 \quad \text{(C–5)}$$

$$E_2(v_k, \sigma_{2k}) \to m_2 = \inf_{v\in\mathcal{A},\sigma_2>0} E_2(v,\sigma_2) \quad \text{(C–6)}$$

$$v_k \rightharpoonup v \text{ weakly in } W^{1,2}(\Omega) \tag{C–7}$$

$$\sigma_{2k} \rightarrow \sigma_2 \tag{C–8}$$

Using the compactness theory, we deduce from (C–4) and (C–7) that

$$u_k \rightarrow u \text{ in } L^2(\Omega), u_k \rightarrow u \text{ a.e. in } \Omega \tag{C–9}$$

$$v_k \rightarrow v \text{ in } L^2(\Omega), v_k \rightarrow v \text{ a.e. in } \Omega \tag{C–10}$$

Since $\{u_k, v_k\} \in \mathcal{A}$, $\{u, v\} \in \mathcal{A}$, thus $E_1(u, \sigma_1) = m_1$ and $E_2(v, \sigma_2) = m_2$, i.e., $\{u, v\}$ is a minimizer for (C–2).

REFERENCES

[1] D. LeBihan and P. J. Basser, "Molecular diffusion and nuclear magnetic resonance," *Diffusion and perfusion magnetic resonance imaging*, pp. 5–17, 1995.

[2] M. E. Moseley, Y. Cohen, J. Mintorovitch, J.suruda L. Chileuitt, D. Norman, and P. Weinstein, "Evidence of anisotropic self-diffusion in cat brain," *Proc. of the 8th Annual Meeting of International Society for Magnetic Resonance in Medicine*, pp. 136–136, 1989.

[3] M. E. Moseley, J. Kucharczyk, H. S. Asgari, and D. Norman, "Anisotropy in diffusion weighted MRI," *Magn. Reson. Med.*, vol. 19, pp. 321–326, 1991.

[4] P.J.Basser and C.Pierpaoli, "Microstructual and physiological features of tissues elucidated by quantitative diffusion tensor mri," *Magn. Reson. Med.*, vol. 111(B), pp. 209–219, 1996.

[5] DS Tuch, RM Weisskoff, JW Belliveau, and VJ Wedeen, "High angular resolution diffusion imaging of the human brain," in *Proc. of the 7th annual meething of International Society for Magnetic Resonance in Medicine (ISMRM'99)*, Philadelphia, 1999, p. 321.

[6] VJ Wedeen, TG Reese, DS Tuchand MR Weigel, J-G Dou, and RM Weisskoffand D. Chesler, "Mapping fiber orientation spectra in cerebral white matter with fourier transform diffusion MRI.," *Proc. of the 8th Annual Meeting of International Society for Magnetic Resonance in Medicine ISMRM'00*, pp. 82–82, 2000.

[7] P. J. Basser, J. Mattiello, and D. Lebihan, "Estimation of the effective self-diffusion tensor from the nmr," *Spin Echo. J. Magn. Reson.*, vol. series B 103, pp. 247–254, 1994.

[8] PJ Basser, J Mattiello, and D LeBihan, "MR diffusion tensor spectroscopy and imaging," *Biophys*, vol. 66:259, pp. 267, 1994.

[9] TL Chenevert, JA Brunberg, and JG Pipe, "Anisotropic diffusion in human white matter: demonstration with MR techniques in vivo," *Radiology*, vol. 177, pp. 401–405, 1990.

[10] EW Hsu and S. Mori, "Analytical expression for the NMR apparent diffusion coefficients in an anisotropy system and a simplified method for determing fiber orientation," *Magn Reson Med*, vol. 34, pp. 194–200, 1995.

[11] Rachid Deriche, David Tschumperle, and Christophe Lenglet, "DT-MRI estimation, regularization and fiber tractography," in *Proc. of 2nd IEEE International Symposium on Biomedical Imaging: From Nano to Macro ISBI'04*, Washington D.C, 2004, pp. 9–12.

[12] P. J. Basser, J. Mattiello, and D. LeBihan, "MR diffusion tensor spectroscopy and imaging," *Biophys*, vol. 66, pp. 259–267, 1994.

[13] A. L. Alexander, K. M. Hasan, M. Lazar, J.S. Tsuruda, and D. L. Parker, "Analysis of partial volume effects in diffusion-tensor MRI," *Magn. Reson. Med.*, vol. 45, pp. 770–780, 2001.

[14] L. Frank, "Characterization of anisotropy in high angular resolution diffusion weighted MRI," in *Proceedings of the 9th Annual Meeting of International Society for Magnetic Resonance in Medicine (ISMRM'01)*, Glasgow, Scotland, 2001, p. 1531.

[15] L. Frank, "Anisotropy in high angular resolution diffusion-weighted MRI," *Magn Reson Med*, vol. 45, pp. 935–939, 2001.

[16] D. S. Tuch, R. M. Weisskoff, J. W. Belliveau, and V. J. Wedeen, "High angular resolution diffusion imaging of the human brain," *Proc. of the 7th Annual Meeting of International Society for Magnetic Resonance in Medicine (ISMRM'99)*, p. 321, 1999.

[17] G. J. M. Parker and D. C. Alexander, "Probabilistic monte carlo based mapping of cerebral connections utilising whole-brain crossing fiber information," in *Information Processing in Medical Imaging*, Ambleside UK, 07 2003, pp. 684–696.

[18] D. S. Tuch, T. G. Reese, M. R. Wiegell, N. Makris, J. W. Belliveau, and V. J. Wedeen, "High angular resolution diffusion imaging reveals intravoxel white matter fiber heterogeneity," *Magn Reson Med*, vol. 48, pp. 577–582, 2002.

[19] Y. Chen, W. Guo, Q. Zeng, Y. Liu, and B.C Vemuri. et. al., "Recovery of intra-voxel structure from high angular resolution diffusion(HARD) MRI," *Proc. of 2nd IEEE International Symposium on Biomedical Imaging: From Nano to Macro (ISBI'04)*, pp. 1028–1031, 2004.

[20] E. O. Stejskal and J. E. Tanner, "Spin diffusion measurements: Spin echoes in the presence of a time-dependent field gradient," *Chem. Phys.*, vol. 42, pp. 288–292, 1965.

[21] D.C.Alexander, G.J.Barker, and S.R.Arridge, "Detection and modeling of non-gaussian apparent diffusion coefficient profiles in human brain data," *Magn.Reson.Med.*, vol. 48, pp. 331–340, 2002.

[22] G. J. M. Parker, J. A. Schnabel, M. R. Symms, D. J. Werring, and G. J. Baker, "Nonlinear smoothing for reduction of systematic and random errors in diffusion tensor imaging," *Magn. Reson. Med.*, vol. 11, pp. 702–710, 2000.

[23] B. C. Vemuri, Y. Chen, M. Rao, T. McGraw, Z. Wang, and T. Mareci, "Fiber tract mapping from diffusion tensor MRI," in *Proc. of IEEE Workshop on Variational and Level Set Methods VLSM'01*, Philadelphia, 2001, pp. 81–88.

[24] C.Feddern, J.Weickert, and B.Burgeth, "Level-set methods for tensor-valued images," in *Proceedings of the 9th Annual Meeting of International Society for Magnetic Resonance in Medicine (ISMRM'03)*, Nice, 2003, pp. 65–72.

[25] C. Poupon, J. F. Mangin, C. A. Clark, V. Frouin, J. Regis, D. LeBihan, and I. Block, "Towards inference of human brain connectivity from MR diffusion tensor data," *Med. Image Anal.*, vol. 5, pp. 1–15, 2001.

[26] D. Tschumperle and R. Deriche, "Regularization of orthonormal vector sets regularization with pde's and applications," *International Journal of Computer Vision*, vol. 50(3), pp. 237–252, 2002.

[27] C. Chefd'hotel, D. Tschumperle', and Olivier D. Faugeras, "Constrained flows of matrix-valued functions: Application to diffusion tensor regularization," in *European Conference on Computer Vision ECCV'02*, 2002, vol. 1, pp. 251–265.

[28] Sinisa Pajevic and Carlo Pierpaoli, "Color schemes to represent the orientation of anisotropic tissues from diffusion tensor data : Application to white matter fiber tract mappying in the human brain," *Magn Reson Med*, vol. 42, pp. 526–540, 1999.

[29] D. Tschumperle and R. Deriche, "Tensor field visualization with pde's and application to DT-MRI fiber visualization," in *Proc. of IEEE Workshop on Variational and Level Set Methods VLSM'03*, Nice, 2003, pp. 256–26.

[30] J. Weickert and T. Brox, "Diffusion and regularization of vector- and matrix-valued images," *Contemporary Mathematics*, vol. 313, pp. 251–268, 2002.

[31] Z. Wang, B. C. Vemuri, Y. Chen, and T. Mareci, "Simultaneous smoothing and estimation of the tensor field from diffusion tensor MRI," in *Proc. of IEEE Conference on Computer Vision and Pattern Recognition CVPR'03*, Wisconxin, 2003, vol. 2.

[32] Z. Wang, B.C.Vemuri, Y.Chen, and T. Mareci, "A constrained variational principle for direct estimation and smoothing of the diffusion tensor field from dwi," in *Proc. of Information Processing in Medical Imaging IPMI'03*, Ambleside, UK, 2003, pp. 660–671.

[33] Y. Chen, W. Guo, Q. Zeng, and Y. Liu, "A nonstandard smoothing in reconstruction of apparent diffusion coefficient profiles from diffusion weighted images," *Inverse Problems and Imaging 2008*, to appear.

[34] Y.Chen, W.Guo, Q.Zeng, and Y.Liu, "Classification of intra-voxel diffusion from HARD MRI," in *Proc. of the 12th annual meething of International Society for Magnetic Resonance in Medicine (ISMRM'04)*, 2004, p. 252.

[35] Y. Chen, W. Guo, Q. Zeng, X. Yan, M. Rao, and Y. Liu, "Apparent diffusion coefficient approximation and diffusion anisotropy characterization in DWI," *19th International Conference on Information Processing in Medical Imaging IPMI'05*, pp. 246–257, 10-15 July 2005.

[36] M. Rao, Y. Chen, B. C. Vemuri, and F. Wang, "Cumulative residual entropy: A new measure of information," *IEEE Trans. on Info. Theory*, vol. 50, pp. 1220–1228, 2004.

[37] Q. Zeng, Y. Chen, W. Guo, and Y. Liu, "Recover multi-tensor structure from HARD MRI under bi-gaussian assumption," *Multiscale Optimization Methods and Applications (Nonconvex Optimization and Its Applications)*, pp. 379–386, 2004.

[38] D. Le Bihan, E. Brethon, and D. Lallemand et.al, "MR imaging of intravoxel incoherent motions: Application to diffusion and perfusion in neurologic disorders," *Radiology*, vol. 161, pp. 401–407, 1986.

[39] B. C. Vemuri, Y. Chen, M. Rao, Z. Wang, T McGraw, T. Mareci, S. J. Blackband, and P. Reier, "Automatic fiber tractography from dti and its validation," in *Proc. of 1st IEEE International Symposium on Biomedical Imaging*, 2002, pp. 505–508.

[40] C. Pierpaoli, P. Jezzard, P. J. Basser, A. Barnett, and G. Di Chiro, "Diffusion tensor MR imaging of the human brain," *Radiology*, vol. 201(3), pp. 637–648, 1996.

[41] C.-F. Westin, S. E. Maier, B. Khidir, P. Everett, F.A. JoleszH., and R. Kikinis, "Image processing for diffusion tensor magnetic resonance imaging," in *Lecture notes in computer science: International Society and Conference Series on Medical Image Computing and Computer-Assisted Intervention MICCAI'99*, Cambridege:Springer, 1999, pp. 441–452.

[42] M.Lazar, D.M.Weinstein, J.S. Tsuruda, Khader M. Hasan, K. Arfanakis, M. Meyerand, B. Badie, H. Rowley, V. Haughton, A. Field, and A. Alexander, "White matter tractography using diffusion tensor deflection," *Human Brain Mapping*, vol. 18, pp. 306–321, 2003.

[43] W. Guo, Q. Zeng, Y. Chen, , and Y. Liu, "Using multiple tensor deflection to reconstruct white matter fiber traces with branching," *3rd IEEE International Symposium on Biomedical Imaging: Macro to Nano(ISBI'06)*, pp. 69–72, 6-9 April 2006.

[44] Q.Zeng, Y.Chen, W.Guo, and Y.Liu, "White matter fiber tracking based on multi-directional vector field," in *Proc. of the 13th annual meething of International Society for Magnetic Resonance in Medicine (ISMRM'05)*, 2005, p. 218.

[45] Q.Zeng, Y.Chen, W.Guo, and Y.Liu, "White matter fiber tracking based on multi-directional vector field," in *Proc. of the 11th Annual Scientific Meeting of the Organization of Human Brain Mapping*, 2005, p. 1649.

[46] D. Hill, P. Batchelor, M. Holden, and D. Hawkes, "Topical review: medical image registration," *Physics in Medicine and Biology*, vol. 46, pp. 1–45, 2001.

[47] B. Zitova and J. Flusser, "Image registration methods: a survey," *Image Vis. comput.*, vol. 21, pp. 977–1000, 2003.

[48] J. Modersitzki, *Numerical methods for image registration*, Oxford, 2004.

[49] J. Maintz and M. Viergever, "A survey of medical image registration," *Medical image analysis*, vol. 2(1), pp. 1–36, 1998.

[50] JPW Pluim, JBA Maintz, and MA Viergever, "Mutual-information-based registration of medical images: a survey," *IEEE Trans. Med. Imaging*, vol. 22, pp. 986–1004, 2003.

[51] Paul A. Viola and William M. Wells III, "Alignment by maximization of mutual information .," in *IEEE International Conference on Computer Vision ICCV'95*, 1995, pp. 16–23.

[52] G. Hermosillo, C.C.Hotel, and O. Faugeras, "Variational methods for multimodal image matching," *Int. J. Computer Vision*, vol. 50(3), pp. 329–343, 2002.

[53] F. Maes, A. Collignon, D. Vandermeulen, G. Marchal, and P. Suetens, "Multi-modality image registration maximization of mutual information," in *MMBIA '96: Proceedings of the 1996 Workshop on Mathematical Methods in Biomedical Image Analysis (MMBIA '96)*, Washington, DC, USA, 1996, p. 14, IEEE Computer Society.

[54] B. Jian, B. Vemuri, and J. Marroquín, "Robust nonrigid multimodal image registration using local frequency maps.," in *19th International Conference on Information Processing in Medical Imaging*, 2005, pp. 504–515.

[55] Ruzena Bajcsy and Stanislav Kovacic, "Multiresolution elastic matching.," *Computer Vision, Graphics, and Image Processing*, vol. 46, no. 1, pp. 1–21, 1989.

[56] J.P. Thirion, "Image matching as a diffusion process: an analogy with maxwell's demons," *Medical image analysis*, vol. 2(3), pp. 243–260, 1998.

[57] Rogelj P and Kovacic S, "Similarity measures for non-rigid registration," *Proc. SPIE*, vol. 4322, pp. 569–78, 2001.

[58] W. Lu, M. Chen, G. Olivera, K. Ruchala, and T. Mackie, "Fast free-form deformable registration via calculus of variations," *Physics in Medicine and Biology*, vol. 49, pp. 3067–3087, 2004.

[59] Coselmon M M, Balter J M, McShan D L, and Kessler M L, "Mutual information based ct registration of the lung at exhale and inhale breathing states using thin-plate splines," *Med. Phys.*, vol. 31, pp. 2942–8, 2004.

[60] Rohlfing T, Maurer C R J, O'Dell W G, and Zhong J, "Modeling liver motion and deformation during the respiratory cycle using intensity-based nonrigid registration of gated mr images," *Med. Phys.*, vol. 31, pp. 427–32, 2004.

[61] H. Wang, L. Dong, J. O'Daniel, R. Mohan, A. Garden, K. Ang, D. Kuban, M. Bonnen, J. Chang, and R. Cheung, "Validation of an accelerated 'demons' algorithm for deformable image registration in radiation therapy," *Phys. Med. Biol.*, vol. 50, no. 12, pp. 2887–2905, June 2005.

[62] Brock K K, Sharpe M B, Dawson L A, Kim S M, and Jaffray D A, "Accuracy of finite element model-based multi-organ deformable image registration," *Med. Phys.*, vol. 32, pp. 1647–59, 2005.

[63] Schreibmann E and Xing L, "Narrow band deformable registration of prostate magnetic resonance imaging, magnetic resonance spectroscopic imaging, and computed tomography studies," *Int. J. Radiat. Oncol. Biol. Phys.*, vol. 62, pp. 595–605, 2005.

[64] B.C. Vemuri, J. Ye, Y. Chen, and C.M. Leonard, "Image registration via level-set motion: Application to atlas-based segmentation," *Medical image analysis*, vol. 7, pp. 1–20, 2003.

[65] N. Ayache, A. Guimond, A. Roche, and J. Meunier, "Three dimensional multimodal brain warping using the demons algorithm and adaptvie intensity correction," *IEEE Trans. Med. Imag.*, vol. 20(1), pp. 58:69, 2001.

[66] R. Bajscy and S. Kovacic, "Multiresolution elastic matching," *Comput. Vision. Graph. Image Process*, vol. 46, pp. 1–12, 1989.

[67] P. Cachier and X. Pennec, "3d non-rigid reigistration by gradient descent on a gaussian-window similarity measure using convolutions," *IEEE workshop on mathematical methods in biomedical image analysis*, pp. 182–189, 2000.

[68] W. Lu, G. H. Olivera, Q. Chen, M Chen, and K. Ruchala, "Automatic re-contouring in 4d radiotherapy," *Physics in Medicine and Biology*, vol. 51, pp. 1077–1099, 2006.

[69] Yan D, Jaffray D A, and Wong J W, "A model to accumulate fractionated dose in a deforming organ," *Int. J. Radiat. Oncol. Biol. Phys.*, vol. 44, pp. 665–75, 1999.

[70] Lu W, Mackie T R, Keller H, Ruchala K J, and Olivera G H, "A generalization of adaptive radiotherapy and the registration of deformable dose distribution," *Proc. 13th Int. Conf. on the Use of Computers in Radiation Therapy (ICCR'00)*, pp. 521–3, 2000.

[71] Olivera G H, Ruchala K, Lu W, Kapatoes J, Reckwerdt P, Jeraj R, and Mackie R, "Evaluation of patient setup and plan optimization strategies based on deformable dose registration," *Int. J. Radiat. Oncol. Biol. Phys.*, vol. 57, pp. S188–9, 2003.

[72] P.J. Keall, J.V. Siebers, S. Joshi, and R. Mohan, "Monte carlo as a four-dimensional radiotherapy treatment-planning tool to account for respiratory motion," *Phys. Med. Biol.*, vol. 49, pp. 3639–3648, 2004.

[73] W. Lu, G.H. Olivera, and T.R.Mackie, "Motion-encoded dose calculation through fluence/sinogram modification," *Med. Phys.*, vol. 32, pp. 118–27, 2005.

[74] P Lorenzen, B Davis, and S Joshi, "Model based symmetric information theoretic large deformation multi-modal image registration," in *IEEE International Symposium on Biomedical Imaging: Macro to Nano, ISBI'04*, 2004, pp. 720 – 723.

[75] G. Christensen and H. Johnson, "Consistent image registration," *IEEE Transactions on Medical Imaging*, vol. 20, pp. 568–582, 2001.

[76] A.D. Leow, S.C. Huang, A. Geng, J. Becker, S. Davis, A. Toga, and P. Thompson, "Inverse consistent mapping in 3d deformable image registration: its construction and statistical properties," *Information Processing in Medical Imaging IPMI'05*, pp. 493–503, 2005.

[77] M.F. Beg and A. Khan, "Symmetric data attachment terms for large deformation image registration," *MedImg*, vol. 26, no. 9, pp. 1179–1189, September 2007.

[78] G. Christensen and H. Johnson, "Invertibility and transitivity analysis for nonrigid image registration," *Journal of Electronic Imaging*, vol. 12, pp. 106–117, jan 2003.

[79] GE Christensen, JH Song, W Lu, Naqa I El, and DA Low, "Tracking lung tissue motion and expansion/compression with inverse consistent image registration and spirometry," *Med Phys*, vol. 34, no. 6, pp. 2155–63, 2007.

[80] Brian Avants and James C. Gee, "Geodesic estimation for large deformation anatomical shape averaging and interpolation," *NeuroImage*, vol. 23 Supplement 1, pp. S139–S150, 2004.

[81] Brian B. Avants, Murray Grossman, and James C. Gee, "Symmetric diffeomorphic image registration: Evaluating automated labeling of elderly and neurodegenerative cortex and frontal lobe," *Biomedical Image Registration*, pp. 50–57, 2006.

[82] Q Zeng and Y Chen, "Accurate inverse consistent non-rigid image registration and its application on automatic re-contouring," *Lecture Notes of Computer Science, proceedings of 4th International Symposium on Bioinformatics Research and Applications(ISBRA'08)*, 2008 to appear.

[83] L. Rudin, S. Osher, and E. Fatemi, "Nonlinear total variation based noise removal algorithm," *Physica D*, vol. 60, pp. 259–268, 1992.

[84] P.Blomgren, T.Chan, P.Mulet, and C.K.Wong, "Total variation image restoration: Numerical methods and extensions," *Proceeding of IEEE Int'l Conference on Image Processing*, vol. 3, pp. 384–387, 1997.

[85] A.Chambolle and P-L.Lions, "Image recovery via total variation minimization and related problems," *Numerische Mathematik*, vol. 1(76), pp. 167–188, 1997.

[86] Y. Chen, S. Levine, and M. Rao, "Variable exponent, linear growth functionals in image restoration," *SIAM Journal on Applied Mathematics*, vol. 66, no. 4(1), pp. 383–406, 2006.

[87] P.Blomgren and T.Chan, "Color tv: total variation methods for restoration of vector-valued images," *IEEE Trans. on Image Processing*, vol. 7(3), pp. 304–309, 1998.

[88] Z. Wang, B. C. Vemuri, Y. Chen, and T. Mareci, "A constrained variational principle for direct estimation and smoothing of the tensor field from complex DWI," *IEEE TMI*, vol. 23:8, pp. 930–939, 2004.

[89] J. Weickert, B. Romeny, and M. Viergever, "Efficient and reliable schemes for nonlinear diffusion filtering," *IEEE Trans. on Img. Proc.*, vol. 7, no. 3, pp. 398–410, March 1998.

[90] T.Lu, P. Neittaanm, and X. Tai, "A parallel splitting up method and its applicaiton to navier-stokes equations," *Applied Mathematics Letters*, vol. 4(2), pp. 25–29, 1991.

[91] T. F. Chan and L. A. Vese, "Active contours without edges," *IEEE Trans. Image Processing*, vol. 10, no. 2, pp. 266–277, 2001.

[92] S.D. Conte and C. DeBoor, *Elementary Numerical Analysis*, McGraw-Hill, New York, 1972.

[93] Y. Chen, W. Guo, Q. Zeng, X. Yan, F. Huang, H. Zhang, G. He, B.C Vemuri, and Y. Liu, "Estimation, smoothing, and charaterization of appranet diffusion coefficient profiles from high angular resolution dwi," *Proc. of IEEE Conference on Computer Vision and Pattern Recognition CVPR'04*, pp. 588–593, 2004.

[94] M. Rao, "More on a new concept of entropy and information," *Journal of Theoretical Probability.*, vol. 18, no. 4, pp. 967–981, 2005.

[95] D. Strong and T. Chan, "Spatial and scale adaptive total variation based regularization and anisotropic diffusion in image processing," *UCLA-CAM Report*, vol. 46, 1996.

[96] DJ Fleet and Y Weiss, "Optical flow estimation," *Mathematical models for Computer Vision: The Handbook, (eds.) N. Paragios, Y. Chen and O. Faugeras*, vol. chapter 15, pp. 239–258, 2005.

[97] E P Simoncelli, E H Adelson, and D J Heeger, "Probability distributions of optical flow," *IEEE Computer Society Conference on Computer Vision and Pattern Recognition,(CVPR'91)*, pp. 310–315, Jun 1991.

[98] Y Weiss and DJ Fleet, "Velocity likelihoods in biological and machine vision," *Probabilistic Models of the Brain: Perception and Neural Function, (eds.) R.P.N. Rao, B.A. Olshausen and M.S. Lewicki*, pp. 81–100, 2001.

[99] GE Christensen, RD Rabbit, and MI Miller, "A deformable neuroanatomy textbook based on viscous fluid mechanics," *Proceedings of the 1993 Conference on Information Sciences and Systems, Johns Hopkins University, March 24-26*, pp. 211–216, 1993.

[100] P.Viola, "Alignment by maximization of mutual information," *PhD thesis*, 1995.

[101] A.Collignon, F.Maes, D.Vandermeulen, P.Suetens, and G. Marchal, "Automated multi-modality image registration based on information theory," *Information Processing in Medical Imaging*, 1995.

[102] G. Hermosillo, "Variational methods for multimodal image registration," *INRIA, France*, 2002.

[103] M Chen, W Lu, Q Chen, KJ Ruchala, and GH Olivera, "A simple fixed-point approach to invert a deformation field," *Med Phys.*, vol. 35, no. 1, pp. 81–88, Jan 2008.

[104] T. Lu, P. Neittaanmki, and X.-C. Tai, "A parallel splitting-up method for partial differential equations and its application to navier-stokes equations," *RAIRO Math. Model. and Numer. Anal*, vol. 26, no. 6, pp. 673–708, 1992.

[105] D. Tuch, "Diffusion MRI of complex tissue structure," *Harvard University and Massachusetts Institute of Technology*, vol. PhD thesis, 2002.

[106] M. Descoteaux, E. Angelino, S. Fitzgibbons, and R.Deriche, "A fast and robust odf estimation algorithm in q-ball imaging," *3rd IEEE International Symposium on Biomedical Imaging: Macro to Nano(ISBI'06)*, pp. 81–84, 6-9 April 2006.

[107] T McGraw, BC Vemuri, B Yezierski, and T Mareci, "von Mises-Fisher mixture model of the diffusion ODF," *3rd IEEE International Symposium on Biomedical Imaging: Macro to Nano(ISBI'06)*, pp. 65–68, 6-9 April 2006.

[108] G.E. Andrews, R. Askey, and R. Roy, *Special Functions*, Cambridge University Press, 1999.

[109] Bing Jian and Baba C. Vemuri, "Multi-fiber reconstruction from diffusion mri using mixture of wisharts and sparse deconvolution," in *Proc. of Information Processing in Medical Imaging IPMI'07*, 2007, pp. 384–395.

[110] J.S.W. Campbell, P Savadjiev, K Siddiqi, and GB Pike, "Validation and regularization in diffusion MRI tractography," *3rd IEEE International Symposium on Biomedical Imaging: Macro to Nano(ISBI'06)*, pp. 351–354, 6-9 April 2006.

[111] R Deriche and M Descoteaux, "Splitting tracking through crossing fibers: multidirectional q-ball tracking," *4th IEEE International Symposium on Biomedical Imaging: Macro to Nano(ISBI'07)*, pp. 756–759, 12-15 April 2007.

[112] JP Byrne, PE Undrill, and RP Phillips, "Feature based image registration using parallel computing methods," *Proceedings of the First Conference on Visualization in Biomedical Computing*, pp. 304–310, May 1990.

[113] X Geng, D Kumar, and GE Christensen, "Transitive inverse-consistent manifold registration," *Inf Process Med Imaging IPMI'05*, pp. 468–479, 2005.

[114] A Jarc, P Rogelj, and S Kovacic, "Texture feature based image registration," *4th IEEE International Symposium on Biomedical Imaging: From Nano to Macro ISBI'07*, pp. 17–20, April 2007.

[115] N. Gupta, "A VLSI architecture for image registration in real time," *IEEE Transactions on Very Large Scale Integration (VLSI) Systems*, vol. 9, no. 15, pp. 981–989, 2007.

[116] R. Strzodka, M. Droske, and M. Rumpf, "Fast image registration in DX9 graphics hardware," *Journal of Medical Informatics and Technologies*, vol. 6, pp. 43–49, Nov 2003.

[117] H Zhang, PA Yushkevicha, and JC Geea, "Registration of diffusion tensor images," *Proceedings of the 2004 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'04)*, vol. 1, pp. I–842 – I–847, 2004.

[118] H Zhang, PA Yushkevicha, DC Alexanderb, and JC Geea, "Deformable registration of diffusion tensor MR images with explicit orientation optimization," *Medical Image Analysis*, vol. 10, no. 5, pp. 764–785, 2006.

[119] W Guo, Y Chen, and Q Zeng, "A geometric flow based approach for diffusion tensor image segmentation," *Philosophical Transactions A*, to appear.

[120] L Zhukov, K Museth, D Breen AH Barr, and R Whitaker, "Level set segmentation and modeling of DT-MRI human brain data," *J. Elec Imag.*, vol. 12, pp. 125–133, 2003.

BIOGRAPHICAL SKETCH

Qingguo Zeng was born in Xiangyang, Hubei Province, P. R. China in 1978. He got his Bachelor of Science degree in 1999 and Master of Engineering in 2002 from Beijing Normal University, Beijing, China. He earned his Ph.D. degree in Applied Mathematics from University of Florida in May 2008. His research interests include mathematical modeling and algorithm developing in medical image processing, statistical image processing, parallel computing, optimization, numerical analysis and numerical Partial Differential Equations.

Invited Paper

The role of Cobalt-60 in modern radiation therapy: Dose delivery and image guidance

L. John Schreiner[1,2], Chandra P. Joshi[1,2], Johnson Darko[1,2], Andrew Kerr[1,2], Greg Salomons[1,2], Sandeep Dhanesar[2]

[1]Department of Medical Physics, Cancer Centre of Southeastern Ontario (CCSEO), Kingston, ON, Canada; [2]Departments of Oncology and Physics, Queen's University, Kingston, ON, Canada

Received on: 23.01.09 Review completed on: 23.03.09 Accepted on: 11.04.09 DOI: 10.4103/0971-6203.54845

ABSTRACT

The advances in modern radiation therapy with techniques such as intensity-modulated radiation therapy and image-guided radiation therapy (IMRT and IGRT) have been limited almost exclusively to linear accelerators. Investigations of modern Cobalt-60 (Co-60) radiation delivery in the context of IMRT and IGRT have been very sparse, and have been limited mainly to computer-modeling and treatment-planning exercises. In this paper, we report on the results of experiments using a tomotherapy benchtop apparatus attached to a conventional Co-60 unit. We show that conformal dose delivery is possible and also that Co-60 can be used as the radiation source in megavoltage computed tomography imaging. These results complement our modeling studies of Co-60 tomotherapy and provide a strong motivation for continuing development of modern Cobalt-60 treatment devices.

Key words: Cobalt-60, intensity-modulated radiation therapy, image guidance, megavoltage computed tomography, tomotherapy

Introduction

Modern radiation therapy is moving towards advanced conformal techniques such as intensity-modulated radiation therapy (IMRT) in conjunction with image guidance, bringing about image-guided radiation therapy (IGRT), to ensure accurate patient treatment. The clinical application of these advanced techniques has been limited almost exclusively to linear accelerators. Investigations of conformal Cobalt-60 (Co-60) radiation delivery have been sparse, in part because of preconceived notions that the radiation beams from Co-60 do not have the properties required in modern radiation treatment.[1] There have been a number of modeling studies that have suggested that Co-60 may be more effective in modern radiation therapy than perceived in the past.[2-4] However, aside from a limited number of studies, there has been little experimental validation of Co-60 conformal delivery.[5] Furthermore, while computed tomography imaging with Co-60 has been performed for basic medical physics studies,[6,7] there has been very little work done on the evaluation of the potential for image guidance with Co-60.

This report will review the results of experimental investigations of the potential of Co-60–based IGRT via tomotherapy — a rotational implementation of IMRT.[8,5] Measured conformal dose distributions achieved with an in-house Co-60 tomotherapy benchtop apparatus will be compared to the corresponding treatment plans. The results of investigations of Co-60 megavoltage computed tomography (MVCT) for image guidance[5,6] will also be shown. The findings of this work support the fact that there is ample potential for administering modern radiation therapy with a cobalt unit and encourage further investigations and development.

Materials and Methods

All irradiations were performed on a Co-60 T780C unit (Best-Theratronics, Kanata, Canada) modified by the addition of a purpose-built computer-controlled rotate translate benchtop apparatus, as shown in Figure 1a. This first-generation tomotherapy test bed includes a rotate translate stage, which can move a phantom to be irradiated through a 1×1 cm^2 Co-60 pencil beam. Intensity-modulated radiation therapy is performed in a single slice

Address for correspondence:
Dr. L. John Schreiner, Department of Medical Physics, Cancer Centre of Southeastern Ontario at Kingston General Hospital, 25 King Street West, Kingston, Ontario K7L 5P9, Canada. E-mail: john.schreiner@krcc.on.ca

by varying the velocity of the phantom as it moves through the pencil beam. By this simple approach, we are well able to imitate the beam delivery from a NOMOS MIMiC multileaf collimator (BEST nomos, Pittsburg, PA, USA, formerly NOMOS Corp., Swickley, PA). It is possible to extend this delivery to 3D by adding a translate capability to the height of the table. Quantitative 3D delivery has been achieved in our laboratory, but the quantitative work reported here will be limited to two dimensions only. The test irradiations were performed on a cylindrical phantom containing GafChromic film (International Specialty Products, NJ, USA). Treatment planning was performed using an in-house developed inverse treatment planning system[9,10] that used either measured pencil beam data in an empirical Milan-Bentley–type algorithm[11] (simple phantom geometries) or EGSnrc/BEAMnrc[12] and EGSnrc/DOSXYZnrc[13] Monte Carlo simulations to model the beam delivery in patient CT data.

The in-house tomotherapy apparatus described above can also be used to perform fan-beam and cone-beam MVCT measurements. Our initial MVCT investigations were based on first-generation (using a pencil beam and diode detector with the image phantom translated through the beam to achieve one projection) and second-generation (using a fan beam and scanning diode) prototype Co-60 tomotherapy imaging systems. Recently our imaging studies have advanced with the development of a third-generation imaging system using a Varian PortalVision LC250 scanning liquid ionization chamber (SLIC) electronic portal imaging device (EPID) (Varian Medical Systems, Palo Alto, CA), shown in Figure 1b. In the Co-60 MVCT experiments, the phantom to be imaged was rotated on the central turntable; typically 180 and 451 image projections were taken through 360 degrees for the diode and SLIC experiments, respectively. The image reconstruction was based on parallel-beam filtered back-projection for diode-based imaging and a generalized Feldkamp technique[14] for EPID fan-beam imaging and was performed using in-house software written in Matlab (The Mathworks, Natick, MA) platform. Anthropomorphic head phantoms, a variety of orthopedic implants, and QA phantoms containing plugs of known electron density were used for qualitative and quantitative assessments.

Results

The findings to date confirm the viability of Co-60–based tomotherapy for conformal dose delivery. Figure 2 shows two examples of conformal irradiations planned for the delivery with the in-house cylindrical treatment phantom — a ring pattern [Figure 2a] and a standard conformal avoidance 'C' plan [Figure 2b]. These plans were implemented to determine whether the inverse treatment planning system would be able to generate an optimized and accurate dose distribution for simple but challenging geometry, including the ability to protect the central critical structure from unwanted radiation.

The ring in Figure 2a represents the PTV region and was prescribed an arbitrary dose of 200 cGy. The dose to the OAR (central structure) was limited to less than 60 cGy. The rest of the volume represents normal tissue (commonly known as the external region), and the dose to this region was restricted to less than 200 cGy. Using the in-house inverse treatment planning program based on aggressive active set (AAS) conjugate gradient method,[9] the plan was generated using 50 field orientations where each orientation used 31 beamlets to provide intensity modulation. The number of beamlets in our system represents the individual pencil beams being projected by the 31 leaf pairs of a multileaf collimator. Therefore, our configuration for field orientations and beamlets is similar to that typically used in commercial tomotherapy. The middle figure in

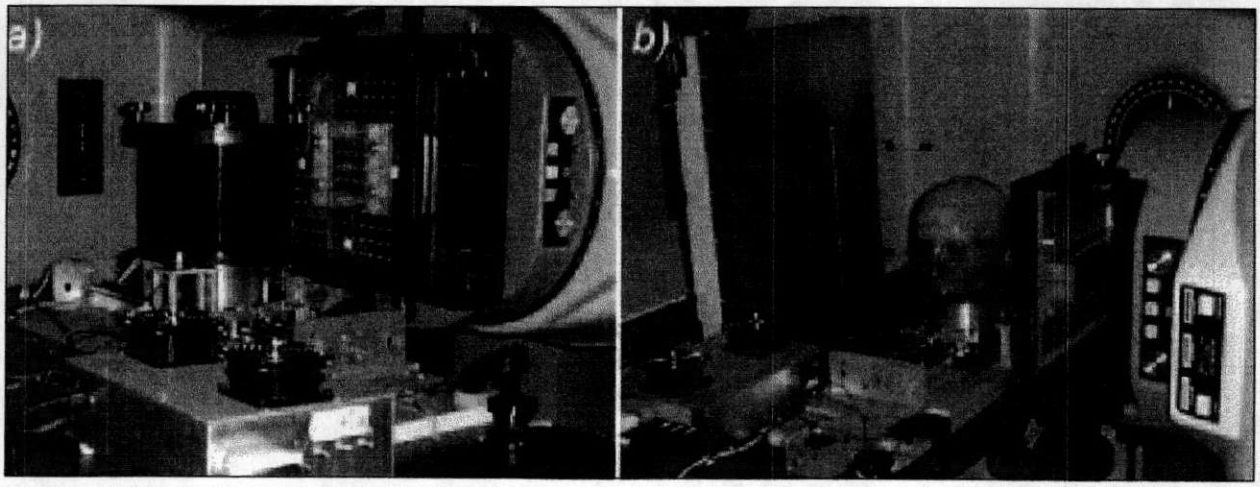


Figure 1: A photograph of the Theratronics T780 Co-60 unit incorporating the tomotherapy benchtop prototype for a) conformal irradiations on film and b) an imaging experiment with a Varian PortalVision LC250 EPID imager

Figure 2a shows the GafChromic film irradiation, which resembles the plan reasonably well. The bottom figure in Figure 2a shows a gamma map[15] of the simulated- and delivered-dose distributions. A gamma value less than or equal to 1 suggests that the agreement between delivery and simulation is within the specified criteria[15]: a standard clinical criterion allows a difference of up to 3% in dose and 3 mm in distance. The gamma comparison of the ring pattern showed that 87% of the treatment volume satisfied the 3%, 3 mm gamma criteria. In clinical practice, the goal is to have 95% of the volume satisfy the gamma criteria. Further analysis showed that 95% of the volume satisfies the gamma criteria if the criteria are 5% and 5 mm, and this is clearly illustrated by the gamma map shown in Figure 2a.

Figure 2b shows a similar plan of an exaggerated concave 'C' structure (the PTV) with an OAR at the center. This plan was generated using a similar number of beam orientations with a dose of 250 cGy to the PTV region and a maximum dose of 100 cGy to the OAR. The delivery was successful, as shown by the middle figure in Figure 2b. The gamma map analysis indicated that the plan and the delivery were in agreement with each other when using the gamma criteria of 5% dose difference and 5 mm distance difference and satisfied the 95% volume requirement.

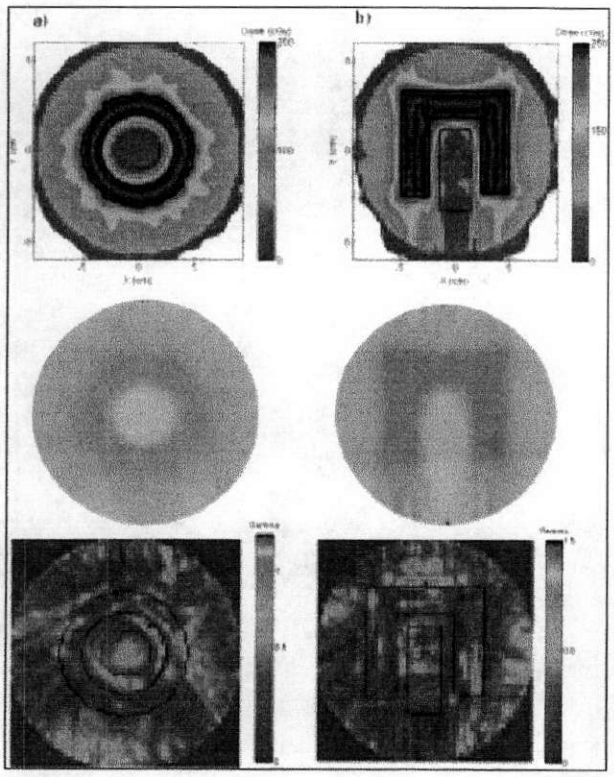


Figure 2: Cobalt-60–based tomotherapy plans of a "ring"-shaped structure (a) and an exaggerated concave "C" structure (b). Top: The dose distributions obtained with our in-house inverse treatment planning system; middle: the deliveries of both patterns on GafChromic film; bottom: gamma maps of the plans and deliveries using 5%, 5 mm gamma criteria

Various treatment plans of similar complexity have been achieved and validated with the GafChromic film measurements.[16] These include simple conformal tests such as a star pattern, a circle, a 'K' pattern and some sophisticated experiments such as the common avoidance 'C' test and a simple head-and-neck test. All measurements showed good agreement with the corresponding plans, indicating that it is possible to deliver Co-60–based tomotherapy irradiations.

Imaging studies have been performed on a number of anthropomorphic phantoms using both a first-generation approach and a third-generation approach using the Varian PortalVision LC250 EPID imager [Figure 1b]. The resulting images illustrate that the Co-60 MVCT scanner performance is, 3 mm high-contrast spatial resolution and 2.8% low-contrast sensitivity (2.5-cm diameter objects). The image in Figure 3a shows one of the early torso images taken using our first-generation approach with a translated/rotated pencil beam and diode detector. Figure 3b shows a third-generation image of a head-and-neck phantom containing three stainless steel pins. The top right image in Figure 3b is of the same phantom obtained with kVCT and clearly shows the image-degrading artifacts associated with the metal pins. These artifacts are absent in the MVCT images shown below in Figure 3. Gross features prominent in the MVCT image include the occipital bone, posteria fosa, temporal bone, maxillary antrum, zygomatic bone, air hole for TLD, and nasal septum.

Discussion

In this paper, we demonstrate that Cobalt-60–based

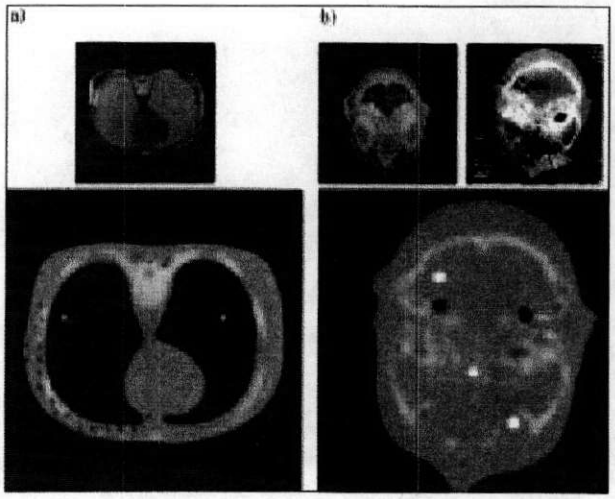


Figure 3: a) Photograph (top) and first-generation Co-60 CT image (bottom) of one torso slice of an anthropomorphic phantom; b) Photograph (top left), kVCT (top right) and third-generation Co-60 CT image (bottom) of one base-of-brain slice of the phantom

tomotherapy approach is capable of both highly conformal intensity-modulated dose deliveries and viable image guidance. While it may seem that the addition of more complex beam collimation and megavoltage imaging to cobalt may complicate a simple technology, and perhaps make the device too complicated or expensive for worldwide distribution, we believe that it is important to establish full modern radiation therapy with a cobalt unit. It is quite clear that in the coming decade, the standard of care for radiation therapy will require image guidance and that radiation units incapable of imaging at the time of delivery will no longer be used in clinics. The development of tomotherapy type approaches with Co-60 also indicates that the potential for use of more sophisticated machines as proposed by Kron *et al.* incorporating MRI[17] may also translate into actual practice. We believe that there are considerable clinical and economic advantages in further investigating modern delivery with Co-60.

Conclusions

Our experiments have confirmed the results of our previous modeling studies that Cobalt-60–based tomotherapy is clinically viable. Two-dimensional conformal dose delivery in a single slice with Co-60 has been validated experimentally using radiochromic film dosimetry. We have successfully developed Co-60 MVCT imaging techniques which provide adequate image quality required for IGRT. We believe that these results clearly indicate that undertaking the task of further development and implementation of modern radiation techniques based on Co-60 units is fully warranted.

Acknowledgments

We thank the Canadian Institute of Health Research (CIHR), Ontario Research Development Challenge Fund (OCITS/ORDCF), Kingston General Hospital (KGH), and the CCSEO for research funding; and NOMOS and MDS Nordion for equipment support. The authors would like to thank Chris Peters for his Co-60–CT assistance with the SLIC imager.

References

1. VanDyk J and Battista JJ. Cobalt-60: An old modality, a renewed challenge. Current Oncology 1996;3:8-17.
2. Cadman P. An investigation of beam parameters for Co-60 tomotherapy. Med Phys 2007;34:3838-43.
3. Fox C, Romeijn HE, Lynch B, Men C, Aleman DM, Dempsey JF. Comparative analysis of ^{60}Co intensity-modulated radiation therapy. Phys Med Biol 2008;53:3175-88.
4. Joshi CP, Darko J, Vidyasagar PB, Schreiner LJ. Investigation of an efficient source design for Cobalt-60 based tomotherapy using EGSnrc Monte Carlo simulations. Phys Med Biol 2008;53:575-92.
5. Schreiner LJ, Kerr AT, Salomons GJ, Dyck C and Hajdok G. The Potential for Image Guided Radiation Therapy with Cobalt-60 Tomotherapy. Proc. of 6th Annual on Medical Image Computing and Computer Assisted Intervention, Springer; Heidelberg, Germany: 2003. p. 449-56.
6. Rogers MV, Kerr AT, Salomons GJ and Schreiner LJ. Quantitative Investigations of Megavoltage Computed Tomography. Proc. of SPIE: Physics of Medical Imaging, M.J. Flynn (ed) International Society for Optical Engineering: Bellingham, WA, 5745; 2005. p. 685-94.
7. Monajemi TT, Tu D, Fallone BG, Rathee S. A bench-top megavoltage fan-beam CT using $CdWO_4$-photodiode detectors. II. Image performance evaluation. Med Phys 2006;33:1090-100.
8. Mackie TR, Holmes T, Swerdloff S, Reckwerdt P, Deasy JO, Yang J, *et al.* Tomotherapy: A new concept for the delivery of conformal radiotherapy. Med Phys 1993;20:1709-19.
9. Ching N, Kerr AT, Rogers M., Schreiner LJ. Development of Inverse Planning and Limited Angle CT Reconstruction for Cobalt-60 Tomotherapy. Proc. of 51st Annual Scientific Meeting of Canadian Organization of Medical Physicists (COMP) and Canadian College of Physicist in Medicine (CCPM), Hamilton, Canada (Hamilton, ON: COMP and CCPM), 2005. p. 235-37.
10. Dhanesar SK, M.Sc. Thesis, An Investigation of Megavoltage Computed Tomography Using a Radioactive Cobalt-60 Gamma Ray Source for Radiation Therapy Treatment Verification, Queen's University, Kingston, ON, 2002.
11. Milan J, Bentley RE. The Storage and manipulation of radiation dose data in a small digital computer. Br J Radiol 1974;47:115-21.
12. Rogers DW, Faddegon BA, Ding GX, Ma CM, We J, Mackie TR. BEAM: A Monte Carlo code to simulate radiotherapy treatment units. Med Phys 1995;22:503-24.
13. Walters B, Kawrakow I and Rogers DWO. DOSXYZnrc Users Manual NRCC Report PIRS-794revB, Ionizing Radiation Standards, National Research Council of Canada, Ottawa, 2005.
14. Feldkamp LA, Davis LC, and Kress JW. Practical cone beam algorithm. J Opt Soc Amer. A1, 612-619, 1984.
15. Low DA, Harms WB, Mutic S, Purdy JA. A technique for the quantitative evaluation of dose distributions. Med Phys 1998;25:656-61.
16. Dhanesar SK, Ching N, Rogers M, Joshi CP, Darko J, Salomons G, Kerr A, Schreiner LJ. Experimental Validation of a Conjugant Gradient based Inverse Treatment Planning System for Cobalt-60 Tomotherapy. Radiother Oncol 84:S21;2007.
17. Kron T, Eyles D, Schreiner LJ, Battista J. Magnetic resonance imaging for adaptive cobalt tomotherapy: A Proposal. J Med Phys 2006;31:242-52.

Source of Support: Canadian Institute of Health Research (CIHR), Ontario Research Development Challenge Fund (OCITS/ORDCF), Kingston General Hospital (KGH), and the CCSEO for research funding. **Conflict of Interest:** None declared.

The British Journal of Radiology, 81 (2008), 304–310

A comparison between cobalt and linear accelerator-based treatment plans for conformal and intensity-modulated radiotherapy

E J ADAMS, MSc and A P WARRINGTON, MSc

Joint Department of Physics, The Royal Marsden NHS Foundation Trust and Institute of Cancer Research, Downs Road, Sutton, Surrey SM2 5PT, UK

ABSTRACT. The simplicity of cobalt units gives them the advantage of reduced maintenance, running costs and downtime when compared with linear accelerators. However, treatments carried out on such units are typically limited to simple techniques. This study has explored the use of cobalt beams for conformal and intensity-modulated radiotherapy (IMRT). Six patients, covering a range of treatment sites, were planned using both X-ray photons (6/10 MV) and cobalt-60 gamma rays (1.17 and 1.33 MeV). A range of conformal and IMRT techniques were considered, as appropriate. Conformal plans created using cobalt beams for small breast, meningioma and parotid cases were found to compare well with those created using X-ray photons. By using additional fields, acceptable conformal plans were also created for oesophagus and prostate cases. IMRT plans were found to be of comparable quality for meningioma, parotid and thyroid cases on the basis of dose–volume histogram analysis. We conclude that it is possible to plan high-quality radical radiotherapy treatments for cobalt units. A well-designed beam blocking/compensation system would be required to enable a practical and efficient alternative to multileaf collimator (MLC)-based linac treatments to be offered. If cobalt units were to have such features incorporated into them, they could offer considerable benefits to the radiotherapy community.

Received 17 May 2007
Revised 26 July 2007
Accepted 9 August 2007

DOI: 10.1259/bjr/77023750

With the advance of increasingly complex treatment techniques in radiotherapy, there is a tendency to move towards ever more sophisticated and expensive external beam delivery equipment. The use of conformal radiotherapy (CFRT) is now routine in most centres, with intensity-modulated radiotherapy (IMRT) also becoming more widespread. Both CFRT and IMRT are usually implemented using a linear accelerator fitted with a multileaf collimator (MLC), to create the conformality and fluence variations necessary [1]. However, the complexity of such machines leads to high maintenance costs and significant planned and unplanned downtime [2, 3].

The advantages of cobalt units, with their very low maintenance costs and minimal need for engineering support, are well known to the radiotherapy community. The net gains in reviving a safer generation of such simple, economical machines could be considerable in a world with increasingly stretched health-care resources. The current generation of three-dimensional treatment planning systems, networked to commercial block and compensator cutters, could provide a practical means of delivering high-quality, radical radiotherapy treatments on cobalt units.

A planning study has therefore been carried out to compare cobalt-60 (Co-60) plans, using conformal blocks and/or compensators, with those created using a 6/10 MV linear accelerator and MLC.

Address correspondence to: E J Adams, Medical Physics Department, Box 152, Addenbrooke's Hospital, Hills Rd, Cambridge CB2 2QQ, UK. E-mail: liz.adams@addenbrookes.nhs.uk

Methods and materials

Planning comparisons were carried out for seven clinical sites using both photons (6/10 MV) and Co-60. Cobalt plans used data from a Theratron Elite cobalt unit (MDS Nordion, Kanata, Canada) and photon plans used data from an Elekta SL15 dual energy accelerator (Elekta, Crawley, UK). The clinical sites were chosen to cover a range of plan complexity, from conventional planning to IMRT, using divergent blocks and compensators for the cobalt unit and a MLC for the accelerator. Planning was carried out on a Helax-TMS system (v6.0.2) (Nucletron UK, Cheshire, UK) using CT scans from previously treated patients. Dose distributions were calculated using a pencil beam calculation algorithm [4]. Conformal plans were normalized such that the planning target volume (PTV) doses were within 95–107% of the prescription dose, ideally with a median dose of 100% [5]. IMRT optimizations also aimed to fulfil these goals although a minimum dose of 90% was considered acceptable provided that 95% of the PTV was enclosed within the 95% isodose; if necessary, IMRT plans were normalized after optimization to achieve this. For the PTVs, minimum and maximum doses (defined as the dose received by 99% and 1% of the volume respectively) were recorded, as were the median dose and the volume covered by the 95% isodose (V95%). For organs at risk

(OARs), dose–volume parameters appropriate for the particular organ, based on the tolerance values, were recorded.

Breast

A patient with right-sided breast cancer was considered for this comparison. The patient separation at the base of the breast was 17 cm, representing a small breast [6], and the maximum lung depth encompassed by the standard tangential fields (measured in beam's eye view) was 1.1 cm. The whole of the breast, excluding the 5 mm beneath the surface of the skin, was contoured as the PTV. Plans were created for Co-60 and 6 MV photons using opposing rectangular wedged tangential fields, angled to give a non-divergent posterior edge. The prescribed dose was 50 Gy.

Oesophagus

The clinical target volume (CTV) for this site comprised the oesophageal tumour with a margin for microscopic spread and the adjacent lymph nodes. This was extended by 3 cm in the superior/inferior direction by contouring, tracking along the oesophagus, and then a 15 mm margin was added circumferentially to create the PTV. Planning was carried out using four conformal fields (gantry angles 0°, 90°, 180°, 270°), shaped using a MLC and blocks for 6 MV and Co-60, respectively. The prescribed dose to the PTV was 54 Gy. The OARs were the spinal cord, to which a maximum dose of 46 Gy was allowed, and the lungs, 20% of which were required to remain below 20 Gy [7].

Meningioma

The CTV in this case was an irregularly shaped base-of-skull meningioma to which a 5 mm margin was added in three dimensions to create the PTV. The prescribed dose was 55 Gy. The brain and brainstem were considered as OARs; although the dose employed was within tolerance limits for these structures [8], functional damage to the brain is both dose and volume dependent and hence it is desirable to limit the volume of normal brain irradiated. Dose to the eyes was minimized by avoiding beam directions that passed through them. CFRT plans were created for Co-60 and 6 MV photons using four and six non-coplanar fields with conformal blocks. Use of conformal blocks for this treatment site is normal clinical practice at our centre because of their superior shaping ability compared with the 1 cm MLC leaves. Additionally, IMRT plans were created for both modalities using the six-field arrangement, using conformal blocks and compensators. In practice, intensity modulation on the accelerator could be achieved using the MLC; however, the planning system would not allow the use of both a MLC and blocks on the same beam.

Parotid

The CTV for this case consisted of the radiographically visible parotid gland surgical bed and the ipsilateral upper cervical lymph nodes, to which a 5 mm margin was applied in three dimensions, excluding the 5 mm beneath the skin surface, to create the PTV. The main OARs under consideration were the ipsilateral cochlea, which lay very close to the PTV, and the contralateral parotid gland. Dose to the eyes was minimized by avoiding beam directions that passed through them, and doses to the oral cavity, brain and brainstem were also considered. Conformal wedged pair and four-field IMRT plans (gantry angles 15°, 40°, 140° and 170° [9]) were created using 6 MV photons with a MLC, and Co-60 with blocks or compensators. The prescribed dose was 60 Gy.

Prostate

The CTV in this case consisted of the prostate plus seminal vesicles, to which a 10 mm margin was added in three dimensions to create the PTV. The prescribed dose was 70 Gy. The main OAR of concern was the rectum, although bladder and femoral heads were also considered; planning goals for the OARs can be seen in Table 1 [10]. A standard three-field conformal plan (gantry angles 0°, 90°, 270°) was prepared using 10 MV photons and a MLC, and then compared with a five-field Co-60 plan (gantry angles 30°, 90°, 180°, 270°, 330°) prepared using conformal blocks. More fields were necessary for the Co-60 plan to keep the doses near the surface of the patient to an acceptable level. Additionally, a six-field non-coplanar Co-60 plan was created, using an anterior and posterior field and lateral fields with the couch twisted by ±30° (the maximum couch twist possible avoiding collision with the treatment head).

Table 1. Planning goals and dose statistics for prostate plans

Organ at risk	Dose (Gy)	Maximum volume (%)	Volume achieved (%)		
			10 MV	5f Co-60	6f Co-60
Bladder	50	50	17.3	14.0	14.9
	60	25	13.7	9.4	10.0
	70	5	3.6	3.4	3.2
Rectum	50	60	36.0	35.3	37.6
	60	50	30.6	29.6	30.6
	65	30	27.3	26.5	25.5
	70	15	12.5	11.6	7.2
Femoral heads	50	50	0.0	0.0	0.0

5f, five-field; 6f, six-field.

Thyroid

For this case the CTV consisted of the thyroid bed and immediately adjacent lymphatics, to which a 5 mm margin was added in three dimensions, excluding the 5 mm beneath the skin surface, to create the PTV. In this situation, where the PTV wraps around the spinal cord, conventional and conformal techniques are unable to deliver the prescribed dose of 60 Gy to the target without exceeding the spinal cord tolerance of 46 Gy [11]; hence, only IMRT plans were considered. Five coplanar fields were used for both Co-60 with compensators and 6 MV photons with a MLC.

Results

Breast

The minimum and maximum PTV doses for the 6 MV plan were 97.2% and 105.3% (median 100.0%) of the prescribed dose respectively; those for the Co-60 plan were 94.5% and 105.7% (median 101.1%) respectively. As the PTV coverage was slightly lower for the Co-60 plan, a second plan was created in which a low-weighted segment was added to the anterior oblique beam with a wedge in the cranio-caudal direction. This second plan had a minimum PTV dose of 95.8% and a maximum of 106.3% (median 101.0%). The volume of lung receiving doses below 25 Gy was slightly increased in the Co-60 plans; however, the lung doses were still perfectly acceptable (volume receiving >20 Gy (V20 Gy) was 2.2% for 6 MV, 3.4% for both Co-60 plans).

Oesophagus

Table 2 lists the dose statistics for the oesophagus plans. Spinal cord and lung doses were both within tolerance; however, the cobalt plan showed poorer target coverage and homogeneity than the 6 MV plan. This was partly due to the large variation in effective path length across the anterior field, where the central part of the beam passes through soft tissue and bone, and the edges pass through a large proportion of lung. This gives rise to large hotspots at the lateral sides of the PTV (see Figure 1a). A second cobalt plan was created, which delivered part of the anterior beam dose through a rectangular segment that included the mediastinum but not the lungs (see Figure 1b). This led to improved target homogeneity as indicated in Table 2.

Meningioma

Table 3 shows the PTV dose statistics for the meningioma plans. In all cases the median dose was 100.0%. The PTV was fully encompassed by the 95% isodose for all conformal plans (V95%=100.0%); for the IMRT plans, V95% was 98.6% for 6 MV and 97.8% for Co-60. Figure 2 shows the dose–volume histograms (DVHs) for the brain and brainstem.

Parotid

Table 4 shows the dose statistics for the parotid case. The minimum PTV dose for the 6 MV IMRT plan was below the goal of 95%; V95% for this plan was 96.4%. The OAR doses were similar between the two modalities. The IMRT plans showed reductions in the doses to the cochlea, oral cavity and contralateral parotid whereas the brainstem doses were increased; however, the brainstem doses were still well within the tolerance of 55 Gy for this structure.

Prostate

Figure 3 shows the DVHs for the PTV, rectum and bladder. The minimum and maximum PTV doses for the 10 MV plan were 97.0% and 105.4% (median 100.0%) of the prescribed dose, respectively; those for the five-field Co-60 plan were 95.0% and 105.3% (median 100.7%), respectively, whereas those for the six-field Co-60 plan were 95.2% and 105.0% (median 100.0%) respectively. Table 1 gives the dose statistics for the OARs compared with the plan acceptance criteria.

Thyroid

Figure 4 shows a transverse slice through the dose distribution for the 6 MV and Co-60 plans. In both cases the isodoses conform tightly around the concavity in the PTV, sparing the spinal cord. Minimum and maximum PTV doses for the 6 MV plan were 92.7% and 105.7% (median 100.7%), respectively, whereas for the Co-60 plan they were 90.0% and 105.7% (median 100.0%), respectively. V95% was 95.0% for the 6 MV plan and 96.2% for the Co-60 plan. The maximum dose to the spinal cord was 38.5 Gy for the 6 MV plan and 46.2 Gy for the Co-60 plan; the doses received by 1 ml of the spinal cord were 36.8 Gy and 38.6 Gy, respectively.

Table 2. Dose statistics for oesophagus plans

		6 MV	Co-60	Co-60 + segment
PTV	Minimum dose (%)	95.2	92.3	93.3
	Maximum dose (%)	106.7	108.7	106.5
	Median dose (%)	99.6	99.6	100.0
	V95% (%)	99.3	94.0	96.4
Spinal cord	Maximum dose (Gy)	40.3	41.9	42.0
Lungs	Volume >20 Gy (%)	13.6	16.2	17.0

PTV, planning target volume.

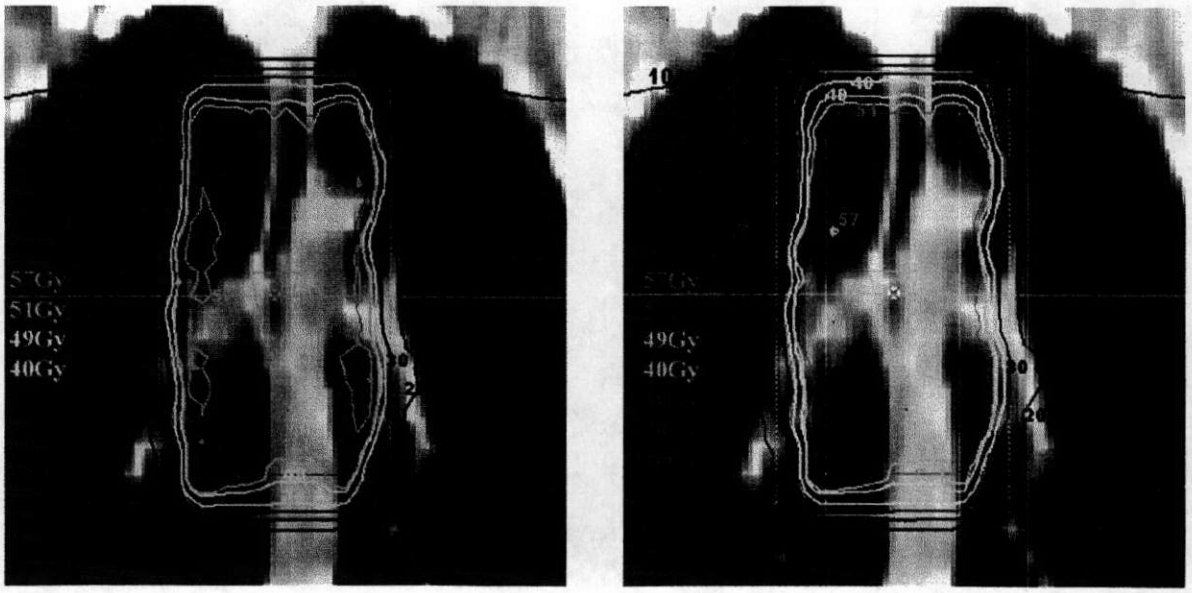


Figure 1. Coronal distribution for Co-60 oesophagus plan (a) without and (b) with an anterior beam segment (shown dashed). Isodoses shown are 10 Gy, 20 Gy, 30 Gy (dark grey), 40 Gy, 49 Gy (light grey), 51 Gy and 57 Gy (mid-grey).

Discussion

As complex radiotherapy techniques become more widespread, there is a tendency for departments to move away from using cobalt units in favour of more sophisticated delivery techniques. However, the simplicity of cobalt units leads to lower maintenance costs and staffing needs as well as exceptional reliability. This paper therefore aimed to investigate the potential of such machines for conformal radiotherapy and IMRT. The treatment sites and plan types were chosen to cover a range of complexity, from simple rectangular fields through to conformal techniques and complex inverse-planned IMRT, including both superficial and deep-seated lesions. Only a single patient case was considered for each site, and hence the results cannot be considered as conclusive; however, they do demonstrate the possibility of designing complex radiotherapy treatments using cobalt.

Plans created for a small-breasted patient showed that cobalt offers a suitable alternative to 6 MV at this site. Although the original cobalt plan did not entirely encompass the PTV with the 95% isodose (minimum dose 94.5%), this was easily rectified by the addition of a wedge in the cranio-caudal direction. This latter cobalt plan showed slightly increased inhomogeneity compared with the 6 MV plan but was within standard International Commission on Radiation Units (ICRU) acceptance criteria. Lung doses were also slightly increased for the cobalt plan because of the larger penumbra for this modality but were still within tolerance criteria.

Two sites were considered to investigate the role of cobalt for conformal planning in the head: meningioma and parotid. In both cases, plans for the two modalities were very similar. The cobalt plans tended to show increased PTV inhomogeneity; for the meningioma case this was still well within acceptable limits and for the parotid the maximum dose only just exceeded the goal of 107% (107.2%) and was therefore considered acceptable. Increasing the number of fields in the meningioma case led to improved OAR sparing at higher dose levels; although the cobalt plans showed slightly increased brain irradiation at some dose levels, these differences were small compared with those caused by changing the number of fields. For the parotid, all OAR doses were comparable. These two treatment sites indicate that cobalt is a suitable modality for planning tumours in the head.

Conformal plans were also created for oesophagus and prostate tumours. For the oesophagus case, the initial cobalt plan showed poor PTV coverage and homogeneity. However, the addition of a rectangular segment to the anterior beam was able to significantly improve the dose distribution. Although the minimum PTV dose was still below the 95% level suggested by the ICRU [5], it was above 93% and more than 95% of the PTV volume was enclosed by the 95% isodose (V95%=96.4%); this

Table 3. Planning target volume (PTV) dose statistics for meningioma plans

	6 MV 4f CFRT	Co-60 4f CFRT	6 MV 6f CFRT	Co-60 6f CFRT	6 MV 6f IMRT	Co-60 6f IMRT
Minimum dose (%)	98.2	97.2	96.6	96.0	94.0	93.3
Maximum dose (%)	102.2	102.9	102.5	102.4	105.4	104.9

4f, four-field; 6f, six-field; CFRT, conformal radiotherapy; IMRT, intensity-modulated radiotherapy.

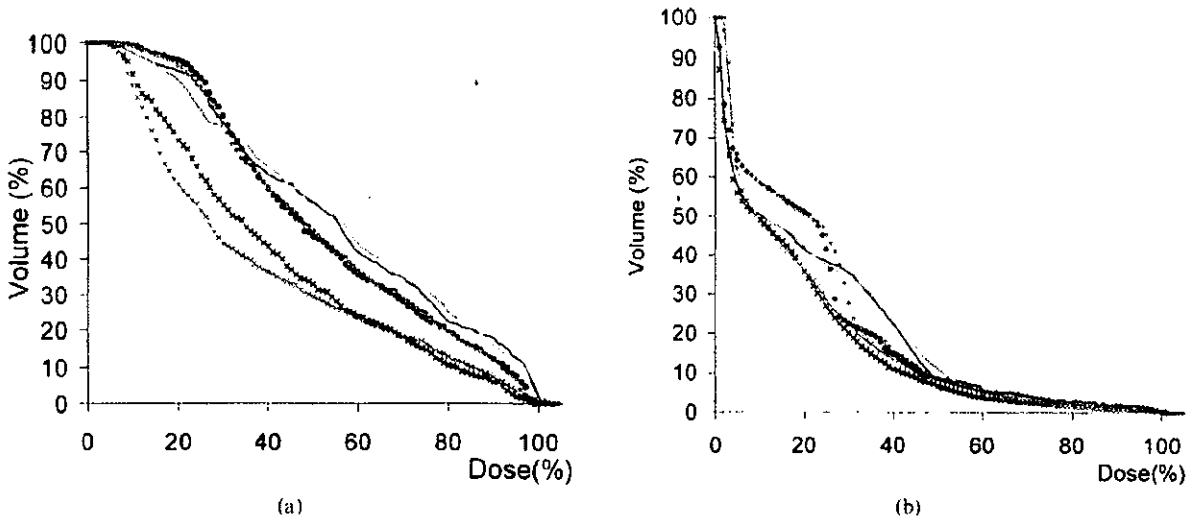


Figure 2. Dose–volume histograms for (a) brainstem and (b) brain from meningioma plans: four-field conformal radiotherapy (CFRT) (solid), six-field CFRT (dotted) and six-field intensity-modulated radiotherapy (IMRT) (crosses) for Co-60 (grey) and 6 MV (black) plans.

would be considered acceptable at our institution. OAR doses were slightly higher for the cobalt plans than for 6 MV, but all were within acceptable limits. This indicates that Co-60 is a suitable modality for planning tumours in the thoracic region. The PTV coverage for the 6 MV plan was better than that for the cobalt plans; however, for this tumour site, the limitations of the pencil beam model used in this study will affect the calculated dose distribution, as neither the lack of scatter from adjacent lung tissue nor the build-up effect for beams passing through lungs will be properly accounted for. In reality, therefore, hotspots will be reduced, as will the coverage of the target in regions where the tumour is next to lung tissue. This latter effect will be more pronounced for 6 MV photons than for Co-60 as the build-up depth is greater, hence making the two modalities more comparable or possibly giving an advantage to the Co-60 plan [12]. A full investigation of this is outside the scope of this study but will be the subject of future work.

Plans using Co-60 were also able to satisfy tolerance criteria for the prostate case; although the PTV coverage was worse for the Co-60 plans than for the 10 MV plan, the minimum dose was above 95% in all cases. The maximum PTV dose was comparable for all plans and all OARs were within the plan acceptance criteria. The bladder doses were lower in the Co-60 plans because of the beam arrangement. Although the five-field Co-60 plan showed higher rectal doses in the 25–50 Gy region, the plans were all very similar for doses above 50 Gy; if doses in the 25–50 Gy region were considered to be important, the six-field Co-60 plan could be used, which gave similar results to those obtained with 10 MV. The need for an increase in the number of fields compared with the 10 MV plan, to maintain normal tissue doses, leads to the cobalt plans being less efficient than the corresponding linac plan. This is particularly true for the non-coplanar six-field plan, although the plan was designed with only two couch positions, thus minimizing the extra effort required. However, this case illustrates that it is possible to treat deep-seated malignancies in the pelvis using Co-60, which have long been cited as being generally unsuitable for this treatment modality.

IMRT planning with cobalt was investigated for parotid, meningioma and thyroid treatment sites. For the parotid case, use of Co-60 IMRT improved both PTV homogeneity and OAR sparing compared with the conformal plan, with the exception of the brainstem, which was still well within tolerance. As for the conformal plans, the differences between the two modalities were small. For the 6 MV IMRT plans, the PTV coverage was reduced compared with the conformal case, with a minimum dose of only 93.1%; however, 96.4% of the PTV was enclosed within the

Table 4. Dose statistics for parotid plans

		6 MV 2f CFRT	Co-60 2f CFRT	6 MV 4f IMRT	Co-60 4f IMRT
PTV	Minimum dose (%)	95.0	95.0	93.1	95.2
	Maximum dose (%)	105.9	107.2	105.8	105.2
	Median dose (%)	100.9	100.0	100.1	100.0
Cochlea	Mean dose (Gy)	38.8	38.8	29.8	30.8
Oral cavity	Mean dose (Gy)	26.3	25.3	19.6	19.0
Contralateral parotid	Mean dose (Gy)	1.7	2.7	1.4	2.2
Brainstem	Maximum dose (Gy)	30.9	28.9	36.8	39.3

2f, two-field; 4f, four-field; CFRT, conformal radiotherapy; IMRT, intensity-modulated radiotherapy, PVT, planning target volume.

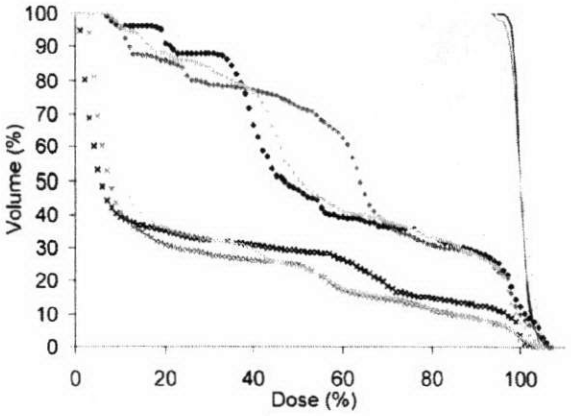
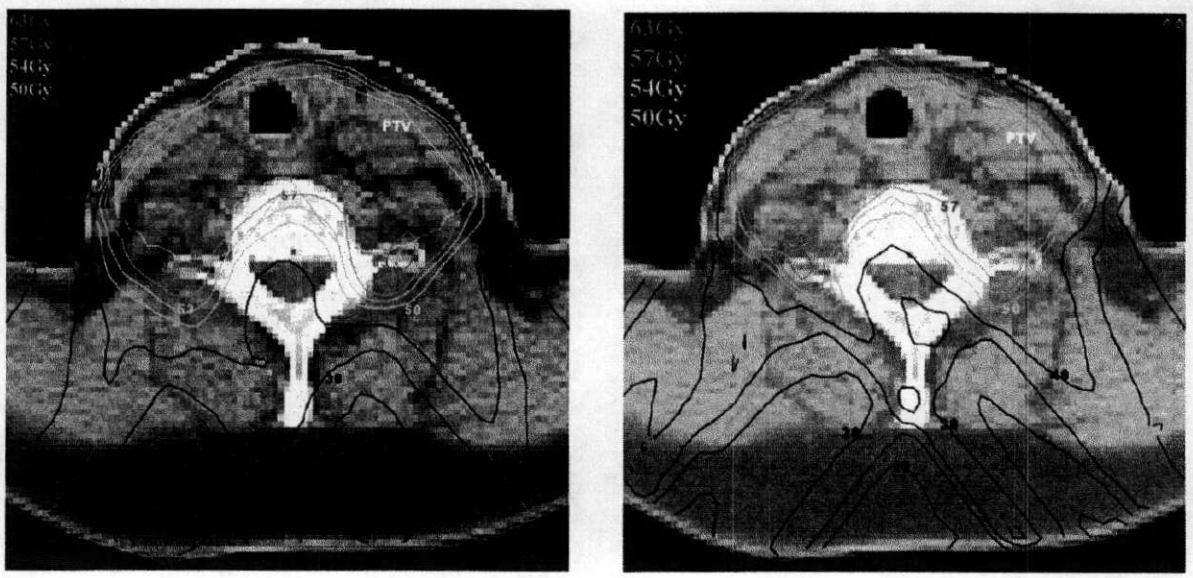


Figure 3. Dose–volume histograms from prostate plans. Planning target volume (PTV) (solid), rectum (dotted) and bladder (crosses) are shown for 10 MV (black), five-field coplanar Co-60 (mid-grey) and six-field non-coplanar Co-60 (light grey) plans.

95% isodose, which is above the criterion of 95% often used for IMRT planning. The meningioma IMRT plans were also comparable between the two modalities, with a significant improvement in OAR doses compared with the conformal plans. Both modalities showed a reduction in PTV coverage compared with the conformal cases, with minimum PTV doses below 95%; however, in both cases V95% was greater than 95% (98.6% for 6 MV, 97.8% for Co-60). The cobalt plan showed a small increase in the volume of brainstem treated to >40 Gy compared with the 6 MV plan, with a significant decrease at lower doses; for the brain, cobalt generally showed a small increase at all dose levels. As for the conformal plans, these differences between modalities were smaller than those between plan types.

For the thyroid case, both modalities were able to cover the PTV with the 90% isodose, with at least 95% of the volume covered by the 95% isodose, whilst sparing the spinal cord. Although the maximum cord dose for the cobalt plan (46.2 Gy) was slightly above the planned constraint of 46 Gy, this dose was only recorded for a very small volume of the cord; the dose received by a more clinically relevant volume (1 mL) was 38.6 Gy.

The Co-60 plans generated in this study used divergent, shaped blocks and/or compensators. Although these beam-shaping devices provide a "gold standard" of conformality, their use in X-ray radiotherapy has been discontinued in many departments because of concerns of overall efficiency, with the MLC providing a more practical alternative. Although it is possible for MLCs to be fitted to cobalt heads, this increase in complexity may compromise the much-cherished simplicity of these machines. Practical CFRT/IMRT delivery on cobalt units would therefore require a well-designed beam blocking/compensation system in order to be an efficient alternative to MLC-based linac treatments. Such a system has been designed by Yoda and Aoki [13], who quote a delivery time of approximately 2 min for a six-field compensated treatment on an accelerator, with approximately 1 min for gantry and collimator rotations, assuming a 2 Gy min^{-1} beam intensity and 2 Gy isocentre dose. Measurements made on a cobalt unit within our department [14], with a 2 Gy min^{-1} source intensity, have suggested IMRT beam delivery times of 20–75 s for a 2 Gy isocentre dose, which, although longer than the times quoted by Yoda and Aoki [13], compare well to linac-based IMRT treatments. A further requirement for precise, radical treatments is the ability to perform verification imaging, preferably electronically; this should be potentially realizable at this energy.



(a) (b)

Figure 4. Transverse dose distributions for (a) 6 MV and (b) Co-60 intensity-modulated radiotherapy (IMRT) thyroid plans. Isodoses shown are 10 Gy, 20 Gy, 30 Gy, 40 Gy (dark grey), 50 Gy, 54 Gy (light grey), 57 Gy and 63 Gy (mid-grey).

Conclusions

Plans of varying complexity, including conformal and IMRT techniques, have been created for Co-60 using blocks and/or compensators for a range of treatment sites. In all cases, the cobalt plans were comparable to those created using 6/10 MV photons with an MLC; in most cases, this was achieved with the same beam arrangement although in some cases, such as the prostate, additional fields were required. These results suggest that it is possible to design high-quality radical radiotherapy treatments using cobalt units. To be able to offer an efficient alternative to MLC-based linac treatments, CFRT/IMRT delivery on cobalt units would require a well-designed beam blocking/compensation system and the ability to perform electronic verification imaging. If cobalt units were to have such features incorporated into them, they could offer considerable benefits to the radiotherapy community.

Acknowledgments

The authors would like to thank the Royal Sussex County Hospital, Brighton, UK for providing the cobalt beam data used in this study, and Dr E M Donovan for helpful comments on the manuscript.

References

1. Board of the Faculty of Clinical Oncology, The Royal College of Radiologists, The Society and The College of Radiographers, Institute of Physics and Engineering in Medicine. Development and implementation of conformal radiotherapy in the United Kingdom. London, UK: The Royal College of Radiologists, 2002.
2. Van der Giesson P-H, Alert J, Badri C, Bistrovic M, Deshpande D, Kardamakis D, et al Multinational assessment of some operational costs of teletherapy. Radiother Oncol 2004;71:347–55.
3. Van der Giesson P-H. Maintenance costs for cobalt machines and linear accelerators: new machines versus old. Letters to the Editor. Radiother Oncol 2002;62:349.
4. Ahnesjo A, Saxner M, Trepp A. A pencil beam model for photon dose calculations. Med Phys 1992;19:263–73.
5. ICRU. Prescribing, recording and reporting photon beam therapy. ICRU Report 62 (Supplement to ICRU Report 50). Bethesda, MD: International Commission on Radiation Units and Measurements, 1999.
6. Venables K, Miles EA, Aird EGA, Hoskin PJ. What is the optimum breast plan – a study based on the START trial plans. Br J Radiol 2006;79:734–9.
7. Graham MV, Purdy JA, Emami B, Harms W, Bosch W, Lockett MA, et al. Clinical dose-volume histogram analysis for pnuemonitis after 3D treatment for non-small cell lung cancer (NSCLC). Int J Radiat Oncol Biol Phys 1999;45:323–9.
8. Emami B, Lyman J, Brown A, Cora L, Goitein M, Munzenrider JE, et al. Tolerance of normal tissue to therapeutic irradiation. Int J Radiat Oncol Biol Phys 1991;21:109–22.
9. Nutting CM, Rowbottom CG, Cosgrove VP, Henk JM, Dearnaley DP, Robinson MH, et al. Optimisation of radiotherapy for carcinoma of the parotid gland: a comparison of conventional, three-dimensional conformal, and intensity-modulated techniques. Radiother Oncol 2001;60:163–72.
10. Adams EJ, Convery DJ, Cosgrove VP, McNair HA, Staffurth JN, Vaarkamp J, et al. Clinical implementation of dynamic and step-and-shoot IMRT to treat prostate cancer with high risk of pelvic node involvement. Radiother Oncol 2004;70:1–10.
11. Nutting CM, Convery DJ, Cosgrove VP, Rowbottom C, Vini L, Harmer C, et al. Improvements in target coverage and reduced spinal cord irradiation using intensity-modulated radiotherapy (IMRT) in patients with carcinoma of the thyroid gland. Radiother Oncol 2001;60:173–80.
12. Knöös T, Ahnesjö A, Nilsson P, Weber L. Limitations of a pencil beam approach to photon dose calculations in lung tissue. Phys Med Biol 1995;40:1411–20.
13. Yoda K, Aoki Y. A multiport compensator system for IMRT delivery. Med Phys 2003;30:880–6.
14. Robinson M. Intensity modulated radiotherapy on a Theratron Elite cobalt unit: investigation of the practical viability of a simple beam modifier for delivering high quality radiotherapy treatments. MSc Thesis. University College London, 2002.

The British Journal of Radiology, 81 (2008), 304–310

A comparison between cobalt and linear accelerator-based treatment plans for conformal and intensity-modulated radiotherapy

E J ADAMS, MSc and A P WARRINGTON, MSc

Joint Department of Physics, The Royal Marsden NHS Foundation Trust and Institute of Cancer Research, Downs Road, Sutton, Surrey SM2 5PT, UK

ABSTRACT. The simplicity of cobalt units gives them the advantage of reduced maintenance, running costs and downtime when compared with linear accelerators. However, treatments carried out on such units are typically limited to simple techniques. This study has explored the use of cobalt beams for conformal and intensity-modulated radiotherapy (IMRT). Six patients, covering a range of treatment sites, were planned using both X-ray photons (6/10 MV) and cobalt-60 gamma rays (1.17 and 1.33 MeV). A range of conformal and IMRT techniques were considered, as appropriate. Conformal plans created using cobalt beams for small breast, meningioma and parotid cases were found to compare well with those created using X-ray photons. By using additional fields, acceptable conformal plans were also created for oesophagus and prostate cases. IMRT plans were found to be of comparable quality for meningioma, parotid and thyroid cases on the basis of dose–volume histogram analysis. We conclude that it is possible to plan high-quality radical radiotherapy treatments for cobalt units. A well-designed beam blocking/compensation system would be required to enable a practical and efficient alternative to multileaf collimator (MLC)-based linac treatments to be offered. If cobalt units were to have such features incorporated into them, they could offer considerable benefits to the radiotherapy community.

Received 17 May 2007
Revised 26 July 2007
Accepted 9 August 2007

DOI: 10.1259/bjr/77023750

With the advance of increasingly complex treatment techniques in radiotherapy, there is a tendency to move towards ever more sophisticated and expensive external beam delivery equipment. The use of conformal radiotherapy (CFRT) is now routine in most centres, with intensity-modulated radiotherapy (IMRT) also becoming more widespread. Both CFRT and IMRT are usually implemented using a linear accelerator fitted with a multileaf collimator (MLC), to create the conformality and fluence variations necessary [1]. However, the complexity of such machines leads to high maintenance costs and significant planned and unplanned downtime [2, 3].

The advantages of cobalt units, with their very low maintenance costs and minimal need for engineering support, are well known to the radiotherapy community. The net gains in reviving a safer generation of such simple, economical machines could be considerable in a world with increasingly stretched health-care resources. The current generation of three-dimensional treatment planning systems, networked to commercial block and compensator cutters, could provide a practical means of delivering high-quality, radical radiotherapy treatments on cobalt units.

A planning study has therefore been carried out to compare cobalt-60 (Co-60) plans, using conformal blocks and/or compensators, with those created using a 6/10 MV linear accelerator and MLC.

Address correspondence to: E J Adams, Medical Physics Department, Box 152, Addenbrooke's Hospital, Hills Rd, Cambridge CB2 2QQ, UK. E-mail: liz.adams@addenbrookes.nhs.uk

Methods and materials

Planning comparisons were carried out for seven clinical sites using both photons (6/10 MV) and Co-60. Cobalt plans used data from a Theratron Elite cobalt unit (MDS Nordion, Kanata, Canada) and photon plans used data from an Elekta SL15 dual energy accelerator (Elekta, Crawley, UK). The clinical sites were chosen to cover a range of plan complexity, from conventional planning to IMRT, using divergent blocks and compensators for the cobalt unit and a MLC for the accelerator. Planning was carried out on a Helax-TMS system (v6.0.2) (Nucletron UK, Cheshire, UK) using CT scans from previously treated patients. Dose distributions were calculated using a pencil beam calculation algorithm [4]. Conformal plans were normalized such that the planning target volume (PTV) doses were within 95–107% of the prescription dose, ideally with a median dose of 100% [5]. IMRT optimizations also aimed to fulfil these goals although a minimum dose of 90% was considered acceptable provided that 95% of the PTV was enclosed within the 95% isodose; if necessary, IMRT plans were normalized after optimization to achieve this. For the PTVs, minimum and maximum doses (defined as the dose received by 99% and 1% of the volume respectively) were recorded, as were the median dose and the volume covered by the 95% isodose (V95%). For organs at risk

(OARs), dose–volume parameters appropriate for the particular organ, based on the tolerance values, were recorded.

Breast

A patient with right-sided breast cancer was considered for this comparison. The patient separation at the base of the breast was 17 cm, representing a small breast [6], and the maximum lung depth encompassed by the standard tangential fields (measured in beam's eye view) was 1.1 cm. The whole of the breast, excluding the 5 mm beneath the surface of the skin, was contoured as the PTV. Plans were created for Co-60 and 6 MV photons using opposing rectangular wedged tangential fields, angled to give a non-divergent posterior edge. The prescribed dose was 50 Gy.

Oesophagus

The clinical target volume (CTV) for this site comprised the oesophageal tumour with a margin for microscopic spread and the adjacent lymph nodes. This was extended by 3 cm in the superior/inferior direction by contouring, tracking along the oesophagus, and then a 15 mm margin was added circumferentially to create the PTV. Planning was carried out using four conformal fields (gantry angles 0°, 90°, 180°, 270°), shaped using a MLC and blocks for 6 MV and Co-60, respectively. The prescribed dose to the PTV was 54 Gy. The OARs were the spinal cord, to which a maximum dose of 46 Gy was allowed, and the lungs, 20% of which were required to remain below 20 Gy [7].

Meningioma

The CTV in this case was an irregularly shaped base-of-skull meningioma to which a 5 mm margin was added in three dimensions to create the PTV. The prescribed dose was 55 Gy. The brain and brainstem were considered as OARs; although the dose employed was within tolerance limits for these structures [8], functional damage to the brain is both dose and volume dependent and hence it is desirable to limit the volume of normal brain irradiated. Dose to the eyes was minimized by avoiding beam directions that passed through them. CFRT plans were created for Co-60 and 6 MV photons using four and six non-coplanar fields with conformal blocks. Use of conformal blocks for this treatment site is normal clinical practice at our centre because of their superior shaping ability compared with the 1 cm MLC leaves. Additionally, IMRT plans were created for both modalities using the six-field arrangement, using conformal blocks and compensators. In practice, intensity modulation on the accelerator could be achieved using the MLC; however, the planning system would not allow the use of both a MLC and blocks on the same beam.

Parotid

The CTV for this case consisted of the radiographically visible parotid gland surgical bed and the ipsilateral upper cervical lymph nodes, to which a 5 mm margin was applied in three dimensions, excluding the 5 mm beneath the skin surface, to create the PTV. The main OARs under consideration were the ipsilateral cochlea, which lay very close to the PTV, and the contralateral parotid gland. Dose to the eyes was minimized by avoiding beam directions that passed through them, and doses to the oral cavity, brain and brainstem were also considered. Conformal wedged pair and four-field IMRT plans (gantry angles 15°, 40°, 140° and 170° [9]) were created using 6 MV photons with a MLC, and Co-60 with blocks or compensators. The prescribed dose was 60 Gy.

Prostate

The CTV in this case consisted of the prostate plus seminal vesicles, to which a 10 mm margin was added in three dimensions to create the PTV. The prescribed dose was 70 Gy. The main OAR of concern was the rectum, although bladder and femoral heads were also considered; planning goals for the OARs can be seen in Table 1 [10]. A standard three-field conformal plan (gantry angles 0°, 90°, 270°) was prepared using 10 MV photons and a MLC, and then compared with a five-field Co-60 plan (gantry angles 30°, 90°, 180°, 270°, 330°) prepared using conformal blocks. More fields were necessary for the Co-60 plan to keep the doses near the surface of the patient to an acceptable level. Additionally, a six-field non-coplanar Co-60 plan was created, using an anterior and posterior field and lateral fields with the couch twisted by $\pm 30°$ (the maximum couch twist possible avoiding collision with the treatment head).

Table 1. Planning goals and dose statistics for prostate plans

Organ at risk	Dose (Gy)	Maximum volume (%)	Volume achieved (%)		
			10 MV	5f Co-60	6f Co-60
Bladder	50	50	17.3	14.0	14.9
	60	25	13.7	9.4	10.0
	70	5	3.6	3.4	3.2
Rectum	50	60	36.0	35.3	37.6
	60	50	30.6	29.6	30.6
	65	30	27.3	26.5	25.5
	70	15	12.5	11.6	7.2
Femoral heads	50	50	0.0	0.0	0.0

5f, five-field; 6f, six-field.

Thyroid

For this case the CTV consisted of the thyroid bed and immediately adjacent lymphatics, to which a 5 mm margin was added in three dimensions, excluding the 5 mm beneath the skin surface, to create the PTV. In this situation, where the PTV wraps around the spinal cord, conventional and conformal techniques are unable to deliver the prescribed dose of 60 Gy to the target without exceeding the spinal cord tolerance of 46 Gy [11]; hence, only IMRT plans were considered. Five coplanar fields were used for both Co-60 with compensators and 6 MV photons with a MLC.

Results

Breast

The minimum and maximum PTV doses for the 6 MV plan were 97.2% and 105.3% (median 100.0%) of the prescribed dose respectively; those for the Co-60 plan were 94.5% and 105.7% (median 101.1%) respectively. As the PTV coverage was slightly lower for the Co-60 plan, a second plan was created in which a low-weighted segment was added to the anterior oblique beam with a wedge in the cranio-caudal direction. This second plan had a minimum PTV dose of 95.8% and a maximum of 106.3% (median 101.0%). The volume of lung receiving doses below 25 Gy was slightly increased in the Co-60 plans; however, the lung doses were still perfectly acceptable (volume receiving >20 Gy (V20 Gy) was 2.2% for 6 MV, 3.4% for both Co-60 plans).

Oesophagus

Table 2 lists the dose statistics for the oesophagus plans. Spinal cord and lung doses were both within tolerance; however, the cobalt plan showed poorer target coverage and homogeneity than the 6 MV plan. This was partly due to the large variation in effective path length across the anterior field, where the central part of the beam passes through soft tissue and bone, and the edges pass through a large proportion of lung. This gives rise to large hotspots at the lateral sides of the PTV (see Figure 1a). A second cobalt plan was created, which delivered part of the anterior beam dose through a rectangular segment that included the mediastinum but not the lungs (see Figure 1b). This led to improved target homogeneity as indicated in Table 2.

Meningioma

Table 3 shows the PTV dose statistics for the meningioma plans. In all cases the median dose was 100.0%. The PTV was fully encompassed by the 95% isodose for all conformal plans (V95%=100.0%); for the IMRT plans, V95% was 98.6% for 6 MV and 97.8% for Co-60. Figure 2 shows the dose–volume histograms (DVHs) for the brain and brainstem.

Parotid

Table 4 shows the dose statistics for the parotid case. The minimum PTV dose for the 6 MV IMRT plan was below the goal of 95%; V95% for this plan was 96.4%. The OAR doses were similar between the two modalities. The IMRT plans showed reductions in the doses to the cochlea, oral cavity and contralateral parotid whereas the brainstem doses were increased; however, the brainstem doses were still well within the tolerance of 55 Gy for this structure.

Prostate

Figure 3 shows the DVHs for the PTV, rectum and bladder. The minimum and maximum PTV doses for the 10 MV plan were 97.0% and 105.4% (median 100.0%) of the prescribed dose, respectively; those for the five-field Co-60 plan were 95.0% and 105.3% (median 100.7%), respectively, whereas those for the six-field Co-60 plan were 95.2% and 105.0% (median 100.0%) respectively. Table 1 gives the dose statistics for the OARs compared with the plan acceptance criteria.

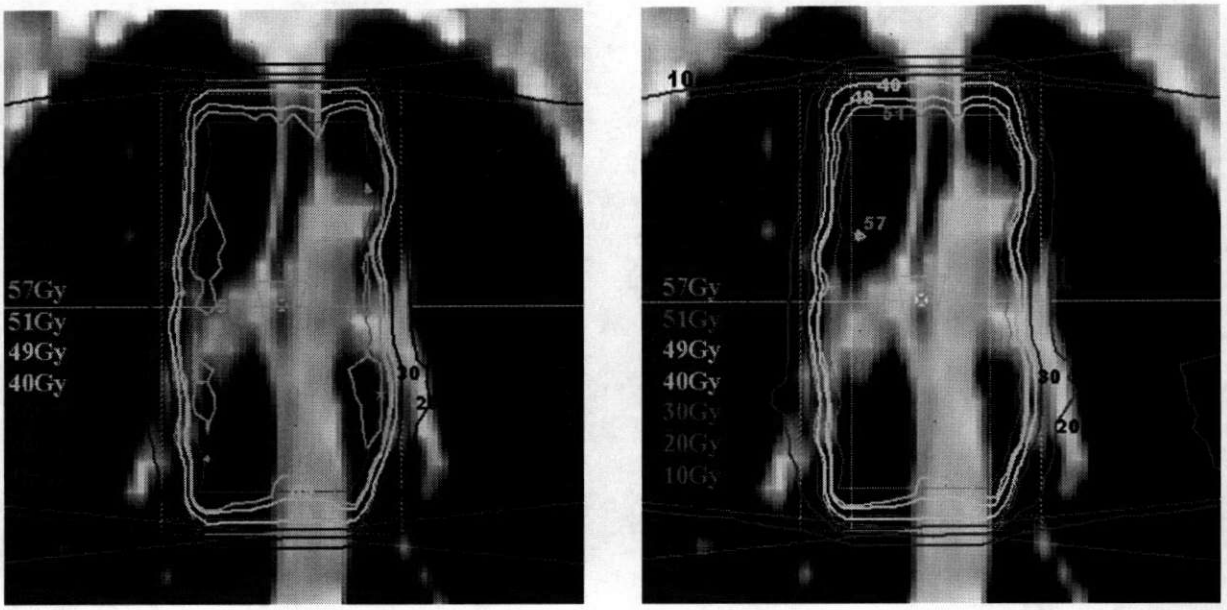
Thyroid

Figure 4 shows a transverse slice through the dose distribution for the 6 MV and Co-60 plans. In both cases the isodoses conform tightly around the concavity in the PTV, sparing the spinal cord. Minimum and maximum PTV doses for the 6 MV plan were 92.7% and 105.7% (median 100.7%), respectively, whereas for the Co-60 plan they were 90.0% and 105.7% (median 100.0%), respectively. V95% was 95.0% for the 6 MV plan and 96.2% for the Co-60 plan. The maximum dose to the spinal cord was 38.5 Gy for the 6 MV plan and 46.2 Gy for the Co-60 plan; the doses received by 1 ml of the spinal cord were 36.8 Gy and 38.6 Gy, respectively.

Table 2. Dose statistics for oesophagus plans

		6 MV	Co-60	Co-60 + segment
PTV	Minimum dose (%)	95.2	92.3	93.3
	Maximum dose (%)	106.7	108.7	106.5
	Median dose (%)	99.6	99.6	100.0
	V95% (%)	99.3	94.0	96.4
Spinal cord	Maximum dose (Gy)	40.3	41.9	42.0
Lungs	Volume >20 Gy (%)	13.6	16.2	17.0

PTV, planning target volume.



(a) (b)

Figure 1. Coronal distribution for Co-60 oesophagus plan (a) without and (b) with an anterior beam segment (shown dashed). Isodoses shown are 10 Gy, 20 Gy, 30 Gy (dark grey), 40 Gy, 49 Gy (light grey), 51 Gy and 57 Gy (mid-grey).

Discussion

As complex radiotherapy techniques become more widespread, there is a tendency for departments to move away from using cobalt units in favour of more sophisticated delivery techniques. However, the simplicity of cobalt units leads to lower maintenance costs and staffing needs as well as exceptional reliability. This paper therefore aimed to investigate the potential of such machines for conformal radiotherapy and IMRT. The treatment sites and plan types were chosen to cover a range of complexity, from simple rectangular fields through to conformal techniques and complex inverse-planned IMRT, including both superficial and deep-seated lesions. Only a single patient case was considered for each site, and hence the results cannot be considered as conclusive; however, they do demonstrate the possibility of designing complex radiotherapy treatments using cobalt.

Plans created for a small-breasted patient showed that cobalt offers a suitable alternative to 6 MV at this site. Although the original cobalt plan did not entirely encompass the PTV with the 95% isodose (minimum dose 94.5%), this was easily rectified by the addition of a wedge in the cranio-caudal direction. This latter cobalt plan showed slightly increased inhomogeneity compared with the 6 MV plan but was within standard International Commission on Radiation Units (ICRU) acceptance criteria. Lung doses were also slightly increased for the cobalt plan because of the larger penumbra for this modality but were still within tolerance criteria.

Two sites were considered to investigate the role of cobalt for conformal planning in the head: meningioma and parotid. In both cases, plans for the two modalities were very similar. The cobalt plans tended to show increased PTV inhomogeneity; for the meningioma case this was still well within acceptable limits and for the parotid the maximum dose only just exceeded the goal of 107% (107.2%) and was therefore considered acceptable. Increasing the number of fields in the meningioma case led to improved OAR sparing at higher dose levels; although the cobalt plans showed slightly increased brain irradiation at some dose levels, these differences were small compared with those caused by changing the number of fields. For the parotid, all OAR doses were comparable. These two treatment sites indicate that cobalt is a suitable modality for planning tumours in the head.

Conformal plans were also created for oesophagus and prostate tumours. For the oesophagus case, the initial cobalt plan showed poor PTV coverage and homogeneity. However, the addition of a rectangular segment to the anterior beam was able to significantly improve the dose distribution. Although the minimum PTV dose was still below the 95% level suggested by the ICRU [5], it was above 93% and more than 95% of the PTV volume was enclosed by the 95% isodose (V95%=96.4%); this

Table 3. Planning target volume (PTV) dose statistics for meningioma plans

	6 MV 4f CFRT	Co-60 4f CFRT	6 MV 6f CFRT	Co-60 6f CFRT	6 MV 6f IMRT	Co-60 6f IMRT
Minimum dose (%)	98.2	97.2	96.6	96.0	94.0	93.3
Maximum dose (%)	102.2	102.9	102.5	102.4	105.4	104.9

4f, four-field; 6f, six-field; CFRT, conformal radiotherapy; IMRT, intensity-modulated radiotherapy.

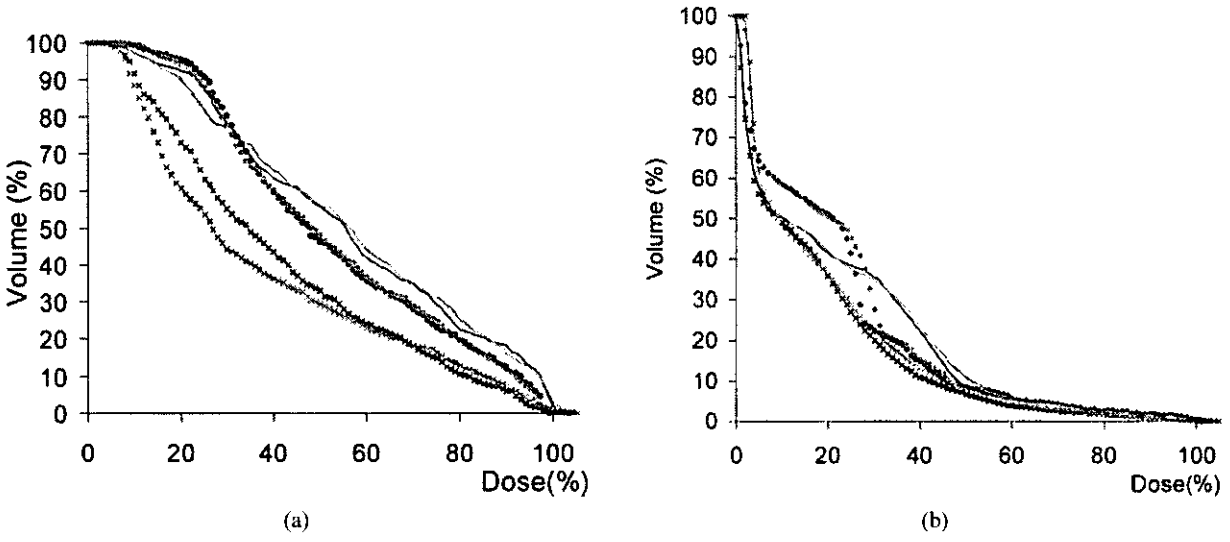


Figure 2. Dose–volume histograms for (a) brainstem and (b) brain from meningioma plans: four-field conformal radiotherapy (CFRT) (solid), six-field CFRT (dotted) and six-field intensity-modulated radiotherapy (IMRT) (crosses) for Co-60 (grey) and 6 MV (black) plans.

would be considered acceptable at our institution. OAR doses were slightly higher for the cobalt plans than for 6 MV, but all were within acceptable limits. This indicates that Co-60 is a suitable modality for planning tumours in the thoracic region. The PTV coverage for the 6 MV plan was better than that for the cobalt plans; however, for this tumour site, the limitations of the pencil beam model used in this study will affect the calculated dose distribution, as neither the lack of scatter from adjacent lung tissue nor the build-up effect for beams passing through lungs will be properly accounted for. In reality, therefore, hotspots will be reduced, as will the coverage of the target in regions where the tumour is next to lung tissue. This latter effect will be more pronounced for 6 MV photons than for Co-60 as the build-up depth is greater, hence making the two modalities more comparable or possibly giving an advantage to the Co-60 plan [12]. A full investigation of this is outside the scope of this study but will be the subject of future work.

Plans using Co-60 were also able to satisfy tolerance criteria for the prostate case; although the PTV coverage was worse for the Co-60 plans than for the 10 MV plan, the minimum dose was above 95% in all cases. The maximum PTV dose was comparable for all plans and all OARs were within the plan acceptance criteria. The bladder doses were lower in the Co-60 plans because of the beam arrangement. Although the five-field Co-60 plan showed higher rectal doses in the 25–50 Gy region, the plans were all very similar for doses above 50 Gy; if doses in the 25–50 Gy region were considered to be important, the six-field Co-60 plan could be used, which gave similar results to those obtained with 10 MV. The need for an increase in the number of fields compared with the 10 MV plan, to maintain normal tissue doses, leads to the cobalt plans being less efficient than the corresponding linac plan. This is particularly true for the non-coplanar six-field plan, although the plan was designed with only two couch positions, thus minimizing the extra effort required. However, this case illustrates that it is possible to treat deep-seated malignancies in the pelvis using Co-60, which have long been cited as being generally unsuitable for this treatment modality.

IMRT planning with cobalt was investigated for parotid, meningioma and thyroid treatment sites. For the parotid case, use of Co-60 IMRT improved both PTV homogeneity and OAR sparing compared with the conformal plan, with the exception of the brainstem, which was still well within tolerance. As for the conformal plans, the differences between the two modalities were small. For the 6 MV IMRT plans, the PTV coverage was reduced compared with the conformal case, with a minimum dose of only 93.1%; however, 96.4% of the PTV was enclosed within the

Table 4. Dose statistics for parotid plans

		6 MV 2f CFRT	Co-60 2f CFRT	6 MV 4f IMRT	Co-60 4f IMRT
PTV	Minimum dose (%)	95.0	95.0	93.1	95.2
	Maximum dose (%)	105.9	107.2	105.8	105.2
	Median dose (%)	100.9	100.0	100.1	100.0
Cochlea	Mean dose (Gy)	38.8	38.8	29.8	30.8
Oral cavity	Mean dose (Gy)	26.3	25.3	19.6	19.0
Contralateral parotid	Mean dose (Gy)	1.7	2.7	1.4	2.2
Brainstem	Maximum dose (Gy)	30.9	28.9	36.8	39.3

2f, two-field; 4f, four-field; CFRT, conformal radiotherapy; IMRT, intensity-modulated radiotherapy, PVT, planning target volume.

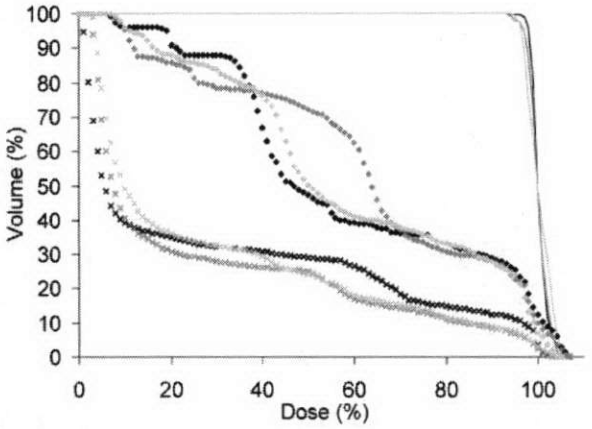
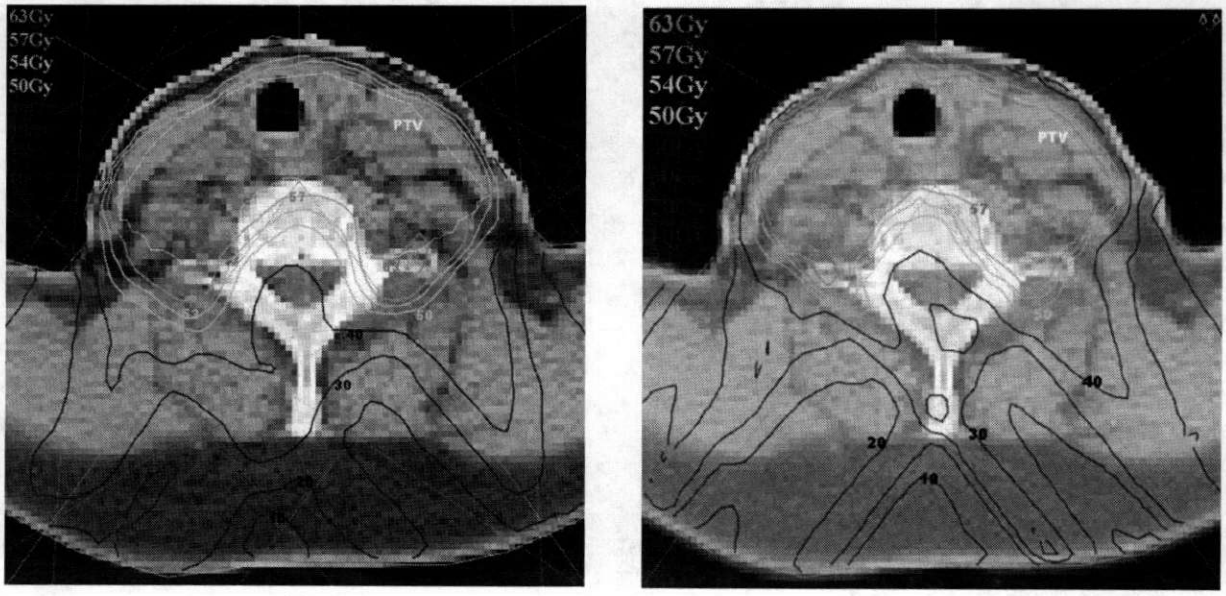


Figure 3. Dose–volume histograms from prostate plans. Planning target volume (PTV) (solid), rectum (dotted) and bladder (crosses) are shown for 10 MV (black), five-field coplanar Co-60 (mid-grey) and six-field non-coplanar Co-60 (light grey) plans.

95% isodose, which is above the criterion of 95% often used for IMRT planning. The meningioma IMRT plans were also comparable between the two modalities, with a significant improvement in OAR doses compared with the conformal plans. Both modalities showed a reduction in PTV coverage compared with the conformal cases, with minimum PTV doses below 95%; however, in both cases V95% was greater than 95% (98.6% for 6 MV, 97.8% for Co-60). The cobalt plan showed a small increase in the volume of brainstem treated to >40 Gy compared with the 6 MV plan, with a significant decrease at lower doses; for the brain, cobalt generally showed a small increase at all dose levels. As for the conformal plans, these differences between modalities were smaller than those between plan types.

For the thyroid case, both modalities were able to cover the PTV with the 90% isodose, with at least 95% of the volume covered by the 95% isodose, whilst sparing the spinal cord. Although the maximum cord dose for the cobalt plan (46.2 Gy) was slightly above the planned constraint of 46 Gy, this dose was only recorded for a very small volume of the cord; the dose received by a more clinically relevant volume (1 mL) was 38.6 Gy.

The Co-60 plans generated in this study used divergent, shaped blocks and/or compensators. Although these beam-shaping devices provide a "gold standard" of conformality, their use in X-ray radiotherapy has been discontinued in many departments because of concerns of overall efficiency, with the MLC providing a more practical alternative. Although it is possible for MLCs to be fitted to cobalt heads, this increase in complexity may compromise the much-cherished simplicity of these machines. Practical CFRT/IMRT delivery on cobalt units would therefore require a well-designed beam blocking/compensation system in order to be an efficient alternative to MLC-based linac treatments. Such a system has been designed by Yoda and Aoki [13], who quote a delivery time of approximately 2 min for a six-field compensated treatment on an accelerator, with approximately 1 min for gantry and collimator rotations, assuming a 2 Gy min^{-1} beam intensity and 2 Gy isocentre dose. Measurements made on a cobalt unit within our department [14], with a 2 Gy min^{-1} source intensity, have suggested IMRT beam delivery times of 20–75 s for a 2 Gy isocentre dose, which, although longer than the times quoted by Yoda and Aoki [13], compare well to linac-based IMRT treatments. A further requirement for precise, radical treatments is the ability to perform verification imaging, preferably electronically; this should be potentially realizable at this energy.



(a) (b)

Figure 4. Transverse dose distributions for (a) 6 MV and (b) Co-60 intensity-modulated radiotherapy (IMRT) thyroid plans. Isodoses shown are 10 Gy, 20 Gy, 30 Gy, 40 Gy (dark grey), 50 Gy, 54 Gy (light grey), 57 Gy and 63 Gy (mid-grey).

Conclusions

Plans of varying complexity, including conformal and IMRT techniques, have been created for Co-60 using blocks and/or compensators for a range of treatment sites. In all cases, the cobalt plans were comparable to those created using 6/10 MV photons with an MLC; in most cases, this was achieved with the same beam arrangement although in some cases, such as the prostate, additional fields were required. These results suggest that it is possible to design high-quality radical radiotherapy treatments using cobalt units. To be able to offer an efficient alternative to MLC-based linac treatments, CFRT/IMRT delivery on cobalt units would require a well-designed beam blocking/compensation system and the ability to perform electronic verification imaging. If cobalt units were to have such features incorporated into them, they could offer considerable benefits to the radiotherapy community.

Acknowledgments

The authors would like to thank the Royal Sussex County Hospital, Brighton, UK for providing the cobalt beam data used in this study, and Dr E M Donovan for helpful comments on the manuscript.

References

1. Board of the Faculty of Clinical Oncology, The Royal College of Radiologists, The Society and The College of Radiographers, Institute of Physics and Engineering in Medicine. Development and implementation of conformal radiotherapy in the United Kingdom. London, UK: The Royal College of Radiologists, 2002.
2. Van der Giesson P-H, Alert J, Badri C, Bistrovic M, Deshpande D, Kardamakis D, et al Multinational assessment of some operational costs of teletherapy. Radiother Oncol 2004;71:347–55.
3. Van der Giesson P-H. Maintenance costs for cobalt machines and linear accelerators: new machines versus old. Letters to the Editor. Radiother Oncol 2002;62:349.
4. Ahnesjo A, Saxner M, Trepp A. A pencil beam model for photon dose calculations. Med Phys 1992;19:263–73.
5. ICRU. Prescribing, recording and reporting photon beam therapy. ICRU Report 62 (Supplement to ICRU Report 50). Bethesda, MD: International Commission on Radiation Units and Measurements, 1999.
6. Venables K, Miles EA, Aird EGA, Hoskin PJ. What is the optimum breast plan – a study based on the START trial plans. Br J Radiol 2006;79:734–9.
7. Graham MV, Purdy JA, Emami B, Harms W, Bosch W, Lockett MA, et al. Clinical dose-volume histogram analysis for pnuemonitis after 3D treatment for non-small cell lung cancer (NSCLC). Int J Radiat Oncol Biol Phys 1999;45:323–9.
8. Emami B, Lyman J, Brown A, Cora L, Goitein M, Munzenrider JE, et al. Tolerance of normal tissue to therapeutic irradiation. Int J Radiat Oncol Biol Phys 1991;21:109–22.
9. Nutting CM, Rowbottom CG, Cosgrove VP, Henk JM, Dearnaley DP, Robinson MH, et al. Optimisation of radiotherapy for carcinoma of the parotid gland: a comparison of conventional, three-dimensional conformal, and intensity-modulated techniques. Radiother Oncol 2001;60:163–72.
10. Adams EJ, Convery DJ, Cosgrove VP, McNair HA, Staffurth JN, Vaarkamp J, et al. Clinical implementation of dynamic and step-and-shoot IMRT to treat prostate cancer with high risk of pelvic node involvement. Radiother Oncol 2004;70:1–10.
11. Nutting CM, Convery DJ, Cosgrove VP, Rowbottom C, Vini L, Harmer C, et al. Improvements in target coverage and reduced spinal cord irradiation using intensity-modulated radiotherapy (IMRT) in patients with carcinoma of the thyroid gland. Radiother Oncol 2001;60:173–80.
12. Knöös T, Ahnesjö A, Nilsson P, Weber L. Limitations of a pencil beam approach to photon dose calculations in lung tissue. Phys Med Biol 1995;40:1411–20.
13. Yoda K, Aoki Y. A multiport compensator system for IMRT delivery. Med Phys 2003;30:880–6.
14. Robinson M. Intensity modulated radiotherapy on a Theratron Elite cobalt unit: investigation of the practical viability of a simple beam modifier for delivering high quality radiotherapy treatments. MSc Thesis. University College London, 2002.

Report of the AAPM Task Group No. 105: Issues associated with clinical implementation of Monte Carlo-based photon and electron external beam treatment planning

Indrin J. Chetty[a)]
University of Michigan, Ann Arbor, Michigan 48109
and University of Nebraska Medical Center, Omaha, Nebraska 68198-7521

Bruce Curran
University of Michigan, Ann Arbor, Michigan 48109

Joanna E. Cygler
Ottawa Hospital Regional Cancer Center, Ottawa, Ontario K1H 1C4, Canada

John J. DeMarco
University of California, Los Angeles, Callifornia 90095

Gary Ezzell
Mayo Clinic Scottsdale, Scottsdale, Arizona 85259

Bruce A. Faddegon
University of California, San Francisco, California 94143

Iwan Kawrakow
National Research Council of Canada, Ottawa, Ontario K1A 0R6, Canada

Paul J. Keall
Stanford University Cancer Center, Stanford, California 94305-5847

Helen Liu
University of Texas MD Anderson Cancer Center, Houston, Texas 77030

C.-M. Charlie Ma
Fox Chase Cancer Center, Philadelphia, Pennsylvania 19111

D. W. O. Rogers
Carleton University, Ottawa, Ontario K1S 5B6, Canada

Jan Seuntjens
McGill University, Montreal, Quebec H3G 1A4, Canada

Daryoush Sheikh-Bagheri
The Regional Cancer Center, Erie, Pennsylvania 16505

Jeffrey V. Siebers
Virginia Commonwealth University, Richmond, Virginia 23298

(Received 11 April 2006; revised 11 July 2007; accepted for publication 18 September 2007; published 27 November 2007)

The Monte Carlo (MC) method has been shown through many research studies to calculate accurate dose distributions for clinical radiotherapy, particularly in heterogeneous patient tissues where the effects of electron transport cannot be accurately handled with conventional, deterministic dose algorithms. Despite its proven accuracy and the potential for improved dose distributions to influence treatment outcomes, the long calculation times previously associated with MC simulation rendered this method impractical for routine clinical treatment planning. However, the development of faster codes optimized for radiotherapy calculations and improvements in computer processor technology have substantially reduced calculation times to, in some instances, within minutes on a single processor. These advances have motivated several major treatment planning system vendors to embark upon the path of MC techniques. Several commercial vendors have already released or are currently in the process of releasing MC algorithms for photon and/or electron beam treatment planning. Consequently, the accessibility and use of MC treatment planning algorithms may well become widespread in the radiotherapy community. With MC simulation, dose is computed stochastically using first principles; this method is therefore quite different from conventional dose algorithms. Issues such as statistical uncertainties, the use of variance reduction techniques, the

ability to account for geometric details in the accelerator treatment head simulation, and other features, are all unique components of a MC treatment planning algorithm. Successful implementation by the clinical physicist of such a system will require an understanding of the basic principles of MC techniques. The purpose of this report, while providing education and review on the use of MC simulation in radiotherapy planning, is to set out, for both users and developers, the salient issues associated with clinical implementation and experimental verification of MC dose algorithms. As the MC method is an emerging technology, this report is not meant to be prescriptive. Rather, it is intended as a preliminary report to review the tenets of the MC method and to provide the framework upon which to build a comprehensive program for commissioning and routine quality assurance of MC-based treatment planning systems. © *2007 American Association of Physicists in Medicine.* [DOI: 10.1118/1.2795842]

Key words: Monte Carlo dose calculation, clinical treatment planning, experimental verification

TABLE OF CONTENTS

I. INTRODUCTION

I.A. Motivation

The accuracy of dose calculations is crucial to the quality of treatment planning and consequently to the doses delivered to patients undergoing radiation therapy.[1] Among other fac-

tors, dose calculations form an integral component in optimizing the therapeutic gain, i.e., maximizing the dose to the tumor for a given normal-tissue dose, for patients treated with radiation. Although the clinical benefit of more accurate dose distributions (i.e., how the improved dose distributions will affect tumor recurrence, i.e., local control, and normal tissue complications) has not been adequately quantified and requires further investigation, evidence exists that dose differences on the order of 7% are clinically detectable.[2] Moreover, several studies have shown that 5% changes in dose can result in 10%–20% changes in tumor control probability (TCP) or up to 20–30% changes in normal tissue complication probabilities (NCTP) if the prescribed dose falls along the steepest region of the dose-effect curves.[3–5] Readers interested in further understanding the need for heterogeneity corrections, among other topics related to dose calculations, are encouraged to read the AAPM Report No. 85,[1] where a comprehensive review of tissue heterogeneity corrections for megavoltage photon beams is provided.

In this report we focus our attention on the Monte Carlo (MC) method, a dose calculation algorithm known to be very accurate when used properly for treatment planning in heterogeneous patient tissues. The issue of lengthy calculation times has traditionally led to the MC method being viewed in the medical physics community as a clinically unfeasible approach. However, the development of MC codes optimized for radiotherapy calculations as well as the availability of much faster and affordable computers, have substantially reduced processing times. These significant advances have led to the clinical use of MC algorithms at some treatment centers and the promised availability of MC photon/electron planning modules among several commercial treatment planning vendors.

In light of the above considerations, MC treatment planning is quickly becoming a reality. An introductory report for the medical physics community on the understanding, implementation, testing, and use of MC algorithms is therefore warranted.

I.B. Objectives for the report

We intend this document to be a preliminary report with the following objectives: (a) to provide an educational review of the physics of the MC method and how it is applied in external beam radiotherapy dose calculations, (b) to describe the role of the MC method in external beam radiotherapy treatment planning process: from the interaction of electrons in the target of the linear accelerator to the deposition of dose in the patient tissues, (c) to describe the issues associated with MC dose calculation within the patient-specific geometry, (d) to discuss the issues associated with experimental verification of MC algorithms, and (e) to discuss the clinical implications of MC calculated dose distributions.

We expect that areas of concern outlined in this report will be further investigated and that more detailed reports providing recommendations on the major issues will be forthcoming.

I.C. Organization of the report

Following the introductory section (Sec. I) we begin in Sec. II with a review of the MC method as it applies to photon and electron transport. We include in this section an overview of MC simulation from the accelerator treatment head to the patient as well as a discussion of variance reduction techniques and efficiency-enhancing methods integral to MC calculations in radiotherapy. Section III begins with a review of the major MC codes being used in clinical application and is followed by detailed discussions on accelerator treatment head modeling and patient-specific treatment planning. This section concludes with the topic of experimental verification, in which guidance is provided on the types of tests needed to verify the accuracy of MC dose calculation algorithms. In Sec. IV we provide a review of recent studies demonstrating the potential clinical impact of MC dose calculations in comparison with conventional algorithms. Finally, we conclude in Sec. V with a summary of the recommendations from this task group report.

II. THE MONTE CARLO METHOD IN RADIOTHERAPY DOSE CALCULATIONS

II.A. Definition of the MC method and historical background

Most generally, the MC technique is a statistical method for performing numerical integrations. MC simulations are employed in many areas of science and technology. Although a method based on random sampling was discussed as early as 1777 by Buffon,[6] the MC technique as we know it today was first developed and named at the end of the second world war. The motivation was to apply MC techniques to radiation transport, specifically for nuclear weapons.[7] The driving forces for the initial idea appear to have been Stanislaw Ulam and John von Neumann who saw the development of ENIAC, the first electronic computer, as an ideal opportunity to develop new applications of statistical sampling. The developments of MC techniques and computers have been closely intertwined ever since, with an exponential increase of the application of MC simulations since digital computers became widely available in the 1950s and 1960s.

Modeling of particle transport problems is ideally suited for the use of MC methods and has been described by Rogers and Bielajew as follows: "The Monte Carlo technique for the simulation of the transport of electrons and photons though bulk media consists of using knowledge of the probability distributions governing the individual interactions of electrons and photons in materials to simulate the random trajectories of individual particles. One keeps track of physical quantities of interest for a large number of histories to provide the required information about the average quantities."[8] As a technique for calculating dose in a patient the underlying physical basis is much simpler in concept than analytic algorithms because the MC method consists of a straightforward simulation of reality and does not involve complex approximations nor models of dose deposition, but only a knowledge of the physics of the various interactions which

have been well understood for over 50 years in most cases. While some of these interactions may be complex to simulate in detail, the basic ideas of each interaction, e.g., an electron giving off a bremsstrahlung photon, are well understood by medical physicists and, hence, the overall process is easy to comprehend.

Although MC was used in several particle physics applications to simulate electron-photon showers in the 1950s, the seminal paper in the field was that of Berger in 1963,[9] in which he described the condensed history technique for electron transport. This technique is the basis of all modern electron-photon transport MC codes relevant to medical physics. The ETRAN code, based on these ideas[10] was developed by Berger and Seltzer and now forms the basis of electron transport in the MCNP code.[11] The release of the EGS4 MC code system in 1985[12] served as a catalyst for the application of the MC method in radiotherapy calculations of dose and dosimeter response. The work of Petti *et al.*,[13] Mohan *et al.*,[14] and Udale[15] being early examples of the use of the EGS MC code system[12,16,17] to simulate medical linear accelerators. Even without the direct use of MC simulations, the MC method already plays a significant role in radiotherapy treatment planning since the energy deposition kernels used in convolution/superposition algorithms have been calculated using MC techniques. Linear accelerator calibration protocols (e.g., AAPM's TG-51)[18] use factors derived from MC simulations. MC-based calculations are also used in the design of treatment head components.[19,20]

Although it has only recently become practical, for over two decades the application of MC techniques to radiation treatment planning has been quite clear.[21,22] The widely used BEAM code system[23] is a pair of EGS4 (now EGSnrc)[17] user codes for simulating radiation transport in accelerators and in patients represented by CT data sets. These relatively easy to use tools have sparked intense research in MC-based radiotherapy treatment planning and have lead to two comprehensive reviews of accelerator simulations by Ma and Jiang[24] and Verhaegen and Seuntjens.[25] Kawrakow and Fippel, among others, have provided the breakthroughs which have made clinical treatment planning feasible, as discussed in Sec. III A. The fast MC codes being developed commercially are almost all based on the results of this collaboration. As this report is being written, the first commercial MC systems have already been introduced into routine clinical treatment planning for electrons[26] and photons.[27]

II.B. Monte Carlo simulation of electron and photon transport

The following material represents a very brief introduction into the MC simulation of electron and photon transport. For more details the reader is referred to the reviews available in the literature.[8,28–31] Another source for detailed information is the documentation accompanying some of the general purpose codes, for instance, the EGSnrc,[17] MCNP,[32] and GEANT4 (Ref. 33) manuals, and the PENELOPE paper.[34]

In the energy range of interest for external beam radiotherapy (megavoltage range), photons interact with surrounding matter via four main processes: incoherent (Compton) scattering with atomic electrons, pair production in the nuclear or electron electromagnetic field, photoelectric absorption, and coherent (Rayleigh) scattering. The first three collision types transfer energy from the photon radiation field to electrons or positrons. In most cases Compton scattering is the dominant interaction, although pair production becomes increasingly important with increasing energy, and may even dominate at higher energies in high-Z components of the treatment head of medical linear accelerators.

When electrons traverse matter, they undergo a large number of elastic interactions and lose energy by two main processes: inelastic collisions with atoms and molecules and radiative interactions. Inelastic collisions result in excitations and ionizations. Ionizations lead to secondary electrons, sometimes referred to as "δ particles". Radiative energy losses, which occur in the form of bremsstrahlung and positron annihilation, transfer energy back to photons and lead to the coupling of the electron and photon radiation fields. One therefore speaks of coupled electron-photon showers.

The electron-photon macroscopic radiation field can be described mathematically by a coupled set of integrodifferential transport equations. These transport equations are prohibitively complicated thereby excluding an analytical treatment except under severe approximations. The MC technique is a solution method that can be applied for any energy range and underlying geometry and material composition.

A solution of the transport problem of particles in matter, which is exact within the existing knowledge of the elementary collision processes, can be obtained by an analog MC simulation. In an analog simulation all particle interactions with surrounding atoms and molecules are explicitly simulated, including those of secondary particles created in the collisions. An analog MC technique is therefore a faithful simulation of physical reality on a digital computer: particles (photons for example) are "born" according to distributions describing the source, they travel a certain distance, determined by a probability distribution, to the site of a collision, and scatter into another energy and/or direction state, possibly creating additional particles. These photons eventually "die" as a result of pair production or photoelectric events or when they Compton scatter to energies below a predetermined low-energy photon cutoff, often called PCUT. Analog simulations, often referred to as "event-by-event" or "interaction-by-interaction" techniques, are typically used for the transport of neutral particles. The analog simulation of charged particle transport is not practical, due to the large number of interactions they undergo until locally absorbed as the energy of the charged particle falls below the predetermined low-energy limit for tracking charged particles, often call ECUT. All general purpose MC codes therefore employ condensed history schemes for charged particle transport, discussed in more detail in Sec. II B 2.

Within a MC simulation, quantities of interest can be computed by averaging over a given set of particle showers (also referred to as "histories," "cases," "trajectories," or "tracks"). One can calculate both observable (measurable)

quantities, such as dose or a particle spectrum, and quantities that cannot easily be measured such as the fraction of particles originating from a certain component of the treatment head, the dose fraction due to scattered photons, etc. Although there are techniques for scoring quantities at a point when using Monte Carlo techniques, in treatment planning applications, it is usual to score quantities (dose mainly) averaged over some finite volume or voxel. As the voxel size is increased, for a given statistical uncertainty the total calculation time will decrease, but the spatial resolution is reduced (see Sec. III D 6 for more discussion). Another important aspect of MC calculations is the presence of statistical uncertainties due to the statistical nature of the method, which is discussed in more detail in Sec. III D 1.

II.B.1. Analog simulations

An analog simulation of particle transport consists of four main steps:

(1) Select the distance to the next interaction.
(2) Transport the particle to the interaction site taking into account geometry constraints.
(3) Select the interaction type.
(4) Simulate the selected interaction.

Steps 1–4 are repeated until the original particle and all secondary particles leave the geometry or are locally absorbed. A particle is considered to be locally absorbed when its energy falls below a specified threshold energy.

Step 1 is based on the probability, $p(r)dr$, that a particle interacts in an interval dr at a distance r from its initial position

$$p(r)dr = e^{-\mu r}\mu dr, \tag{1}$$

where μ is the linear attenuation coefficient (number of interactions per unit length). A random distance r distributed according to $p(r)$ can be sampled using the so-called inverse-transform method, which equates the cumulative probability of $p(r)$ with a random number ξ distributed uniformly between zero and unity

$$\int_0^r p(r')dr' = \xi \Rightarrow r = -\frac{\ln(1-\xi)}{\mu}. \tag{2}$$

Step 2 involves basic ray tracing, which requires a geometry model that can provide the medium and mass density of a region together with a computation of the distance to the next geometry boundary along the particle trajectory.

Step 3 is similar to step 1 except that now the probability distribution function is discrete, i.e., it involves a fixed number of final states i, corresponding to an interaction of type i. Suppose that the cross section for interaction of type i is denoted by σ_i and the total cross section by $\sigma=\Sigma\sigma_i$. A direct application of the inverse-transform method for n interaction types yields interaction *1* if $\xi \leq \sigma_1/\sigma$, else interaction *2* if $\xi \leq (\sigma_1+\sigma_2)/\sigma$, else interaction n, if $\xi \leq (\sigma_1+\sigma_2+\cdots+\sigma_n)/\sigma$.

Perhaps the most difficult part is step 4, where one must sample energy/direction changes from the differential cross section of the selected process. The manuals of the popular general purpose codes[17,32–34] provide details of the methods employed for the relevant photon interactions.

Based on the above discussion it should be clear that an analog MC simulation is conceptually quite straightforward.

II.B.2. Condensed history simulations

The condensed history technique was first described comprehensively in the pioneering work by Berger.[9] It is based on the observation that the vast majority of electron interactions lead to very small changes in the electron energy and/or direction. Many such "small-effect" interactions can therefore be grouped into relatively few condensed history "steps" and their cumulative effect taken into account by sampling energy, direction, and position changes from appropriate distributions of grouped single interactions, e.g., multiple scattering, stopping power, etc. Berger defined two main classes of condensed history implementations. In a class I scheme all collisions are subject to grouping. The effect of secondary particle creation above specified threshold energies are taken into account after the fact (i.e., independently of the energy actually lost during the step) by setting up and transporting the appropriate number of secondary particles. In this way the correlation between large energy losses and secondary particle creation is lost. In a class II scheme interactions are divided into "hard" (sometimes also referred to as "catastrophic") and "soft" collisions. Soft collisions are subject to grouping as in a class I scheme; hard collisions are explicitly simulated in an analog manner.

A class II scheme can be described with the same four basic steps that make up an analog simulation. The two main differences are that only hard collisions are included and that step 2 is much more difficult because the particles do not move on straight lines and because it involves the selection of energy, direction, and position changes from multiple scattering distributions. It is also frequently necessary to divide the distance between catastrophic interactions into shorter condensed history steps to guarantee the accuracy of the simulation. As in an analog simulation there is a transport threshold energy. Particle transport thresholds are often the same as the particle production thresholds dividing hard and soft collisions, but this is not a necessary condition.

A class II MC simulation is illustrated in Fig. 1. The upper portion shows a complete electron track including secondary electrons and photons (shown with dashed lines and not including their interactions) with energies above the hard collision thresholds. The lower portion is a magnified view of the shaded box. The actual curved path has been simulated using four condensed history steps. The filled circles and arrows show the positions and directions at the beginning of the steps. The shaded area around the electron track indicates the region where the energy of subthreshold secondary particles is in reality deposited. If this volume is small compared to the calculation resolution (i.e., voxel size in the case of radiotherapy calculations), energy deposition can be considered local and modeled using a restricted stopping power along the electron track. Note that the initial and final posi-

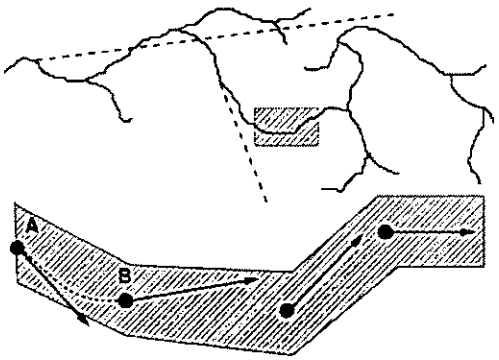


FIG. 1. Illustration of a class II condensed history scheme for electron transport. The upper portion shows a complete electron track including secondary electrons and photons (shown with dashed lines and not including their interactions) with energies above the hard collision thresholds. The lower portion is a magnified view of the shaded box.

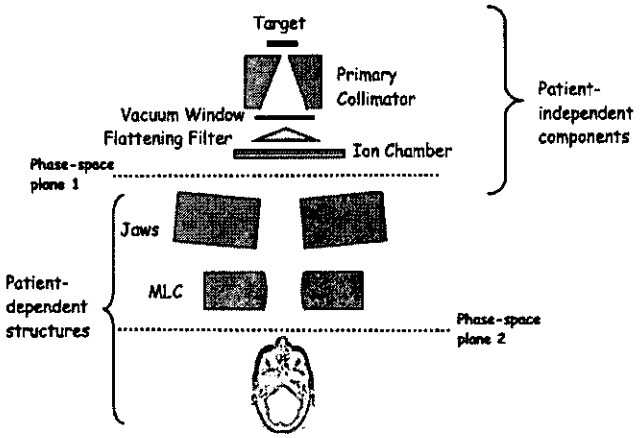


FIG. 2. Illustration of the components of a typical Varian linear accelerator treatment head in photon beam mode. Phase space planes for simulating patient-dependent and patient-independent structures are also represented. For other manufacturers, component structures (such as the jaws, MLC, etc.) may be in different locations, thereby potentially requiring a change in the placement of the phase space scoring planes.

tions of a step are not connected in the figure to highlight the fact that a condensed history implementation does not provide information on how the particle goes from *A* to *B* (the curved dashed line connecting *A* and *B* is a more realistic representation of the trajectory than a straight line from *A* to *B*). This becomes important when the scoring grid is not the same as the underlying geometry grid or when a single condensed history step traverses more than one geometrical region as in some of the fast MC codes specialized for use in radiotherapy.[35–39] This consideration highlights another important aspect of a condensed history simulation, namely the way the transport is performed in the vicinity of or across boundaries between different regions. The EGSnrc code, for instance, utilizes single scattering (i.e., analog) simulation within a certain perpendicular distance from an interface.[40] Although this approach is necessary for accurate simulations of certain types of geometries, it is generally not needed for typical radiotherapy calculations.

Although the condensed history technique makes use of practical MC simulations possible, it introduces the step size as an artificial parameter. Dependencies of the calculated results on the step size have become known as step-size artifacts.[41] Step-size artifacts were a major factor in the early years of most general purpose MC codes. Due to significant theoretical developments in the nineties the condensed history technique is now well understood.[42,43] This has lead to the development of high accuracy condensed history implementations[40,44] and faster MC codes that can compute dose distributions with accuracy comparable to traditional MC packages in a small fraction of the time.[35–39,45]

With charged particle transport one stops tracking the particle's movement at some low-energy cutoff and the choice of the cutoff can affect the calculation in two important ways. The higher the value of the cutoff, the faster the calculation; this can improve the calculation speed significantly. On the other hand, unless great care is taken, stopping at too high a cutoff energy can distort the dose distribution since the "stopped" charged particle might have deposited energy some distance from where its trajectory was terminated. Thus care must be taken in selecting an energy cutoff.

II.C. Overview of Monte Carlo-based radiotherapy dose calculations

It is possible to carry out a single MC simulation in which one starts with the electron exiting from the accelerator structure, follows it and its descendants (e.g., bremsstrahlung photons, knock-on electrons) through the fixed elements of the head (targets or scattering foils, primary collimators, monitor chambers, flattening filters, etc.), the various beam shaping devices which are patient specific (jaws, multileaf collimator (MLC), applicators, cutouts, wedges, compensators), and finally the patient as specified by a CT or some other data set. This tracking of the initial particle and all of its descendants is referred to as a history. As discussed further in Sec. III D 1.1, it is important to include all particles associated with a single initiating electron as part of the same history.

Due to significant improvements in the efficiency of photon beam treatment head simulations,[46] the speed of the complete simulation is such that it is feasible to consider performing the entire calculation for each patient.[47] However, there have been a variety of strategies for dividing such calculations into several steps. The first step, transporting particles through the patient-independent elements, can be inefficient without the use of advanced variance reduction techniques (see Sec. II D). This is especially true for photon beams, since many bremsstrahlung photons generated in the target will strike the primary collimator and not contribute to the beam reaching the patient. One approach to improve the simulation efficiency is to first perform the simulation of the patient-independent structures and to store what is called a phase-space file at a plane just below the fixed elements of the accelerator head (see phase-space plane 1 in Fig. 2). The phase-space file contains phase-space parameters for all particles as they cross the scoring plane. The phase-space parameters consist of the energy, position, direction, charge, and possibly other information such as the region/s of cre-

ation or interaction of particles. The advantage of this approach is that this part of the calculation can be reused as often as necessary. Particles are then transported through the patient specific collimation system and are either stored in another phase-space file at the base of the accelerator (see phase-space plane 2 in Fig. 2) or tracked through the patient in the same simulation. Storing a second phase-space may be more efficient when open fields (e.g., 10×10 cm^2 fields) are used for treatment, however, more commonly, when MLCs are used for beam shaping, the latter approach is likely to be more efficient.

As we will discuss in Sec. III C, the entire phase-space data for the accelerator may also be generated using beam modeling (virtual source model) approaches which do not require direct MC simulation of the accelerator for each treatment. One class of virtual source models is based on characterizing the results of a MC simulation of the accelerator head and another class is based solely on measured beam data such as depth-dose curves, profiles and output ratios. In either case, the patient-dependent components (e.g., the MLC) are simulated using either explicit transport methods or approximate transport methods before detailed transport in the patient.

As with conventional planning methods, experimental verification forms an integral part of the clinical implementation of a MC dose calculation algorithm. MC algorithms will benefit from similar experimental verification procedures as conventional systems, and as such should follow the commissioning procedures detailed in the AAPM TG-53 report[48] and other relevant publications.[49,50] Experimental verification in more complex fields and/or heterogeneous geometries is useful to verify the expected improved accuracy associated with MC calculations in these situations.

II.D. Variance reduction techniques and efficiency enhancing methods

The efficiency, epsilon (ε), of a MC calculation is defined as: $\varepsilon = 1/s^2T$, where s^2 is an estimate of the true variance (σ^2) of the quantity of interest and T is the CPU time required to obtain this variance. Since both Ns^2 and T/N are approximately constant, the efficiency is roughly independent of N, the number of histories simulated. There are two ways to improve the efficiency of a given calculation: either decrease s^2 for a given T or decrease T for a given N while not changing the variance. Techniques which improve the efficiency by changing the variance for a given N while not biasing the result (i.e., not changing the expectation value which is the value expected in an infinitely long run) are called variance reduction techniques. Variance reduction techniques often increase the time to simulate a single history and are only useful if the overall efficiency is improved. A given technique may increase the efficiency for some quantities being scored and decrease it for others. In contrast to variance reduction techniques, there are a variety of ways to speed up a given calculation by making an approximation which may or may not affect the final result in a significant way.

Bremsstrahlung splitting and Russian roulette of secondary particles are widely used variance reduction techniques which are especially useful in simulating an accelerator treatment head.[23,46] In the various forms of bremsstrahlung splitting, each time an electron is about to produce a bremsstrahlung secondary, a large number of secondary photons with lower weights are set in motion, the number possibly depending on a variety of factors related to the likelihood of them being in the field. If the number of photons created is selected to minimize those that are not directed toward the patient plane, then there is a further saving in time. Russian roulette can be played whenever there is little interest in a particle resulting from a specific class of events. The low interest particles are eliminated with a given probability, but to ensure an unbiased result, the weights of the surviving particles are increased by the inverse of that probability. A common example is to play Russian roulette with secondary electrons created from photon interactions in treatment head structures. Another variance reduction method, photon forcing, may sometimes be used to enhance the production of electrons in the air downstream of the accelerator. In a photon forcing scheme, the parent photon is forced to interact in a given geometric region and the weights of the resulting particles are adjusted accordingly to maintain an unbiased result.

Range rejection and increasing the energy at which electron histories are terminated (energy cutoffs) are examples of methods which, when used correctly, improve efficiency by decreasing the time per history without significantly changing the results. In range rejection, an electron's history is terminated whenever its residual range is so short that it cannot escape from the current region or reach the region of interest. In most implementations this ignores the possible creation of bremsstrahlung photons while the electron loses energy which means this is an approximate technique. When applied to electrons below a certain energy threshold, this form of range rejection produces the same results in a reduced computing time.[23] It is also possible to implement range rejection in a manner which properly accounts for bremsstrahlung production and thus make it an unbiased variance reduction technique. By stopping tracking of electrons at a higher energy, efficiency can be improved, but this may have an effect on the dose distribution if too high a threshold is used. Playing Russian roulette with particles at energies below a relatively high transport cutoff or with range-rejected particles is a comparable variance reduction technique for reducing the simulation time. However, its implementation is typically more difficult and this has favored the use of range rejection and high transport cutoffs in situations where it is easy to demonstrate that the resulting error is sufficiently small.

There are other variance reduction and efficiency enhancing techniques which collectively have allowed substantial increases in the speed of a calculation. These methods include the reuse of particle tracks,[37] and other adaptations of particle track reuse, such as the simultaneous transport of particle sets (STOPS) approach of Kawrakow,[36] which is a variance reduction technique. For a more comprehensive re-

view of variance reduction and efficiency enhancing techniques, readers are referred to a chapter by Rogers and Bielajew[51] and the papers by Rogers *et al.*[23] and Kawrakow *et al.*[46,52] Other useful references on the influence of variance reduction methods in the context of phantom and patient treatment planning are provided in the articles by Ma *et al.*[53] and Kawrakow and Fippel.[45]

In summary, variance reduction techniques are an important requirement for the use of MC calculations in the clinical setting; without them, calculation times would still be too long for use in most situations. However, inappropriate use of a variance reduction technique can reduce calculation efficiency, thus increasing calculation time. In principle variance reduction techniques, implemented correctly, do not alter the physics and thereby produce unbiased results. Other efficiency improving techniques can significantly alter the accuracy of the calculation if applied inappropriately. Improper implementations can lead to unpredictable results in either case.

It is incumbent upon the medical physicist to understand, at a minimum, those techniques that the user can adjust in a clinical MC algorithm. In addition, tests must be done to show correct implementation of those techniques over the range of clinical situations. Vendors should provide adequate documentation for users to understand the techniques employed and how the implementation was validated.

III. MONTE CARLO SIMULATION OF RADIATION TRANSPORT IN ACCELERATORS AND PATIENTS

III.A. Review of current Monte Carlo codes

A large number of general purpose MC algorithms have been developed for simulating the transport of electrons and photons. Perhaps the most widely used of these in medical physics is the EGS code system.[12,16,17,40,44] There are several other comparable general purpose systems used in medical physics such as the ITS (Refs. 54 and 55) and MCNP systems[11,32] both of which have incorporated the electron transport algorithms from ETRAN (Refs. 10 and 56) which was developed at NIST by Berger and Seltzer following the condensed history techniques proposed by Berger.[9] Other newer general purpose systems include PENELOPE (Ref. 34) and GEANT4.[33] The EGS and ITS/ETRAN and MCNP systems are roughly of the same efficiency for calculations in very simple geometries when no variance reduction techniques are used, whereas the other systems tend to be considerably slower. An important special purpose code is the EGS user code, BEAM.[23,57–60] The BEAM code is optimized to simulate the treatment head of radiotherapy accelerators and includes a number of variance reduction techniques to enhance the efficiency of the simulation.[46] Comprehensive reviews of MC simulation of radiotherapy beams from linear accelerators are available elsewhere.[24,25]

While the accuracy of these general-purpose codes can be roughly the same as long as they are carefully used, these codes are considered too slow for routine treatment planning purposes. Several groups have published on the use of parallelization of MC techniques over multiple computers to provide more reasonable turn-around times for simulation in clinical research.[61–63] Specific to radiation therapy, there have been a variety of MC codes developed to improve the calculation efficiency, especially in the patient simulation. The PEREGRINE system (North American Scientific: Nomos Division) was developed at the Lawrence Livermore National Laboratory and has been benchmarked against measurements.[27] The PEREGRINE electron transport algorithm is a modified version of the EGS4 condensed history implementation. PEREGRINE uses the random hinge approach[64] for electron transport mechanics. Several efficiency enhancing and variance reduction techniques are implemented in PEREGRINE, including source particle reuse, range rejection, Russian roulette and photon splitting. Parallelizing the calculation on several computer processors is also implemented to reduce the overall dose calculation time. The system decouples the scoring zones from the transport geometry. Source modeling in PEREGRINE is achieved by performing a full MC simulation of the accelerator head using the BEAM code[23] and using the output to create a source model[65] from which source particles are regenerated above the patient-dependent beam modifiers. PEREGRINE uses several approximations when transporting the beam through the patient specific beam modifiers, followed by transport through a patient's CT data set.[66] The PEREGRINE system was the first MC algorithm to receive FDA 510-K approval and represents the first commercially available photon beam treatment planning system in the United States.

Several commercial MC implementations currently available or under development are based on the Voxel Monte Carlo (VMC) series of codes. The initial version[37] of VMC was only applicable to electron beams and involved several approximations in the modeling of the underlying interaction processes. Improved treatment of multiple elastic scattering[67] was incorporated in 1996, PENELOPE's random hinge method in 1997,[68] and all remaining approximations removed in 2000.[45] A photon transport algorithm was added in 1998,[35] which included precalculated interaction densities in each voxel similar to approaches developed previously.[69,70] The resulting code was named XVMC. In 1999, a series of advanced variance reduction techniques were developed and incorporated into XVMC which brought an additional factor of 5–9 increase in simulation speed. Treatment planning applications and experimental verification of VMC-based systems have been reported in several articles.[71–75] Separate versions of the VMC code were subsequently developed by Fippel (XVMC) (Refs. 45 and 76 and Kawrakow (VMC++)[36]. XVMC is being incorporated into the Monaco (CMS), PrecisePlan (Elekta), and *i*Plan (BrainLab) treatment planning systems. VMC++ includes additional refinements in the physics models, such as the exact Kawrakow–Bielajew multiple scattering formalism,[77] including relativistic spin effects,[17] and the STOPS method (mentioned in Sec. II D). VMC++ is the basis for the first commercial electron MC algorithm from Nucletron and is being incorporated into the Masterplan (Nucletron) and Eclipse (Varian) treatment planning systems for photon beam dose calculations. The VMC/XVMC/

VMC++ code systems have also been integrated into several MC-based research systems including those at the University of Tubingen, McGill University, and the Virginia Commonwealth University.

Another MC code that has reached commercial implementation is the Macro Monte Carlo (MMC) method[38,78] for electron beam treatment planning. MMC uses the MC technique, but is very different from the standard simulation of radiation transport. MMC uses a precalculated database from EGSnrc simulations of electron transport through small spheres of varying sizes and materials and follows a random walk through the CT phantom based on these precalculated values. The commercial implementation of MMC, eMC, (Eclipse, Varian) makes use of some precalculated accelerator-specific information; however, fluence intensities arising from the various subsources are fitted to the user's measured data.[79] More details on the performance of the eMC system for clinical electron dose calculations is available elsewhere.[80]

There are several institutions currently engaged in developing MC radiotherapy applications for clinical and/or research related purposes. MCDOSE (Ref. 53 and 81) is among the first of these types of systems. MCDOSE is based on EGS4 and includes fundamental changes in some aspects of the electron transport in order to improve speed. MCDOSE has been shown to give results very similar to EGS4.[53] It performs particle tracking through the beam modifiers in conjunction with the patient calculation and has built-in capability to handle various models of the incident beam. The speed ups have been obtained by using various techniques (bremsstrahlung splitting, photon forcing, track repetition,[37] and range rejection).

The MCV (Monte Carlo Vista) [82] code is used for clinical IMRT planning and verification[61] as well as for a variety of research related applications. MCV interfaces photon-electron MC dose algorithms to the Pinnacle (Philips Radiation Oncology Systems, Madison, WI) commercial planning system, and calculations are performed in a parallel environment using multiple Unix-based processors.[82] Treatment head simulation is accomplished using a modified version of the BEAM code, with calculations divided into two stages, based on the patient-dependent[83] and patient-independent component structures.[82] Patient and phantom calculations (within MCV) are completed using DOSXYZnrc,[58] VMC++ (described above), or MCVRTP, a C++ MC code developed by Philips that uses many of the algorithms of EGS4.[82] MCV utilizes variance reduction techniques inherent to the subcodes it uses, and achieves speed by use of multiple processors.[82]

Another major research code is the dose planning method[39] (DPM) code system developed initially for performing electron beam dose calculations in a voxelized geometry. DPM utilizes the Kawrakow–Bielajew multiple scattering formalism[77] and the random hinge approach for transport mechanics.[64] Particles do not stop at boundaries[39] as is the case with other fast codes. DPM has been integrated into the University of Michigan's in-house treatment planning system (UMPlan) and is currently being used for a variety of photon beam treatment planning studies.[84,85]

An MCNP (Monte Carlo *N*-particle)-based code, RT_MCNP (Ref. 86) has been in use for treatment planning research at UCLA. There have been a series of publications related to the use of RT_MCNP for a variety of applications, from radiosurgery to IMRT planning using a micromultileaf collimator.[86–91] Finally, treatment planning studies using the GEANT,[92–94] PENELOPE,[95] gamma electron positron transport system (GEPTS),[96,97] and ORANGE (Ref. 98) (MCNP-based) MC codes have also been reported.

In an effort to quantify the speed and accuracy for the phantom component of the calculations by the various MC codes being used for research and/or clinical planning purposes, Rogers and Mohan[99] proposed what came to be known as the ICCR benchmark. The tests and geometries for the ICCR benchmark comparisons were as follows:[99] (a) speed test: phantom of dimensions $30.5 \times 30.5 \times 30$ cm with $(5\ \text{mm})^3$ voxels filled either randomly with one of 4 materials (water, aluminum, lung, and graphite) or with water alone, 6 MV photons (spectrum) from a point source at 100 cm SSD and collimated to $10 \times 10\ \text{cm}^2$ at the phantom surface, (b) accuracy test: heterogeneous phantom as defined in (a) with $5 \times 5 \times 2$ mm voxels (2 mm along the depth axis), 18 MV photons (spectrum) from a point source at 100 cm SSD and collimated to $1.5 \times 1.5\ \text{cm}^2$ at the phantom surface. Beam spectra[268] were provided for these comparisons in order to standardize the beam model used in the dose calculations. Statistical uncertainties were to be reported as the relative uncertainty in the dose for voxels with a dose greater than some arbitrary lower limit, such as 50% of the maximum dose.[99] Results for the ICCR benchmark are summarized in Table 1. Some timing results have been added recently. Timing values have been scaled to that on a single Intel P-IV 3.0 GHz processor.

The reader should be aware that the timing results reported in Table I are susceptible to large variations (on the order of at least 20%) due to variations in compilers, memory size, cache, etc.[99] Timing comparisons for more clinically relevant treatment plans are presented in Sec. III E 4. These results have been reported by medical physicists using commercial MC systems for treatment planning.

III.B. Accelerator treatment head simulation

III.B.1. Sensitivity of simulations to electron beam and other parameters

In general one does not know all the details of the clinical accelerator. For example, the characteristics of the incident electron beam are only known approximately. Knowledge of the sensitivity of MC simulation results to input parameters, such as the position, direction, and energy of the initial electron beam exiting the accelerator and to details of the geometry of the treatment head, is important. A sensitivity analysis is indispensable in determining which source and geometry parameters to adjust and by how much in order to improve agreement with user-specific measurements.

Factors influencing the characteristics of a photon beam are the energy, spatial, and angular distributions of the electrons incident on the target (or exiting the waveguide), and

TABLE I. Summary of timing and accuracy results from the ICCR benchmark. Timing comparisons were performed using 6 MV photons, 10×10 cm^2 field size, and those for the accuracy test, using 18 MV photons and a 1.5×1.5 cm^2 field size, as detailed in the ICCR benchmark (Ref. 99). All times have been scaled to the time it would take running on a single, Pentium IV, 3 GHz processor. Readers should be aware that the timing results, as well as the method used to scale the times, are subject to large uncertainties due to differences in compilers, memory size, cache size, etc.

Monte Carlo code	Time estimate (min)	% mean difference relative to ESG4/PRESTA/DOSXYZ
ESG4/PRESTA/DOSXYZ	43	0, benchmark calculation
VMC++	0.9	±1
XVMC	1.1[a]	±1
MCDOSE (modified ESG4/PRESTA)	1.6	±1
MCV (modified ESG4/PRESTA)	22	±1
DPM (modified DPM)	7.3[b]	±1
MCNPX	60[c]	Maximum difference of 8% at Al/lung interface (on average ±1% agreement)
PEREGRINE	43[d]	±1
GEANT4 (4.6.1)	193[e]	±1 for homogeneous water and water/air interfaces

[a]Results not originally part of the ICCR benchmark study, from Ref. 76.
[b]Results not originally part of the ICCR benchmark study, reported independently by author I.J.C. for the modified version of DPM developed for clinical planning calculations.
[c]Timing results not originally part of the ICCR benchmark study, reported independently by author J.J.D. Calculations were performed using the *F8 (energy deposition) tally.
[d]Timing and accuracy results not originally part of the ICCR benchmark study, reported independently by author D.S.-B.
[e]Timing and accuracy results not originally part of the ICCR benchmark study. Estimated independently by author Seuntjens based on Poon and Verhaegen (Ref. 94) and ICCR type speed-test dose calculations (Ref. 270). Timing results are for the standard physics model; the low-energy and the PENELOPE model lead to a factor of 2 more CPU time. The accuracy result reported is derived from the interface perturbation studies in Poon and Verhaegen (Ref. 94), applies to 1.25 MeV monoenergetic photon beams and represents the difference with EGSnrc using PRESTA-II electron step algorithm and "exact" boundary crossing. For a water/Pb interface and 1.25 MeV photons, the maximum difference with EGSnrc increases to 6% (Ref. 94).

the dimensions, materials, and densities of all the components interacting with the beam (the target, primary collimator, flattening filter(s), monitor chamber, collimating devices, such as blocks or MLCs, and beam modifying devices such as wedges). Several investigators have reported on the sensitivity of megavoltage beam simulations to the electron beam striking the target and other treatment head parameters.[19,100–103] Faddegon *et al.*,[19] in simulating Siemens accelerators, showed that the key parameters are the mean energy and focal spot size of the electron beam incident on the exit window, the material composition and thickness profile of the exit window, target, flattening filter, primary collimator, and the position of the primary collimator relative to the target. Bieda *et al.*[101] showed that the accelerator simulation for 20 MeV (Varian-produced) electron beams was very sensitive to the distance between the scattering foils and, to a lesser extent, to the width of the shaped secondary scattering foil. Changes to the primary or secondary foil thickness were found to significantly alter the falloff and bremsstrahlung components of the depth-dose curve.[101] Sheikh-Bagheri and Rogers[102] performed calculations of "in-air" off-axis ratios and depth-dose curves and compared these with measurements to derive estimates for the parameters of the electron beam incident on the target, and to study the effects of some mechanical parameters, such as target width, primary collimator opening, flattening filter material, and density. Their study[102] included several different photon beam energies from accelerators produced by different manufacturers (Varian, Siemens, and Elekta). The electron beam radial intensity distribution was found to influence the off-axis ratios to a great extent. The greater the width of the electron-beam radial intensity distribution, the relatively more intense is the photon beam on the central axis.[102] Figure 3 shows the influence of the electron-on-target energy and radial intensity FWHM on 40×40 cm^2 field profile doses from Tzedakis *et al.*[103] The calculated profiles are observed to be quite sensitive to these parameters. The central axis depth dose curves are also strongly influenced by the electron-on-target energy.[102] However, the central-axis depth-dose curves are quite insensitive to variation in the radial intensity distribution of the electron beam striking the target, because the dose along the central axis is deposited primarily by particles in the vicinity of the central axis. The divergence of the electron beam incident on the target also needs to be considered as it may affect large field profiles.

Regarding the influence of individual treatment head components, Sheikh–Bagheri and Rogers[102] (see Table II) showed that even small changes (0.01 cm) in the primary collimator's upstream opening can affect in-air off-axis ratios by restricting the number of bremsstrahlung photons contributing to the scattered photon fluence reaching off-axis points downstream. Dose profiles are quite sensitive to the composition and density of the flattening filter as noted in Fig. 4, where two different flattening filter materials were used for

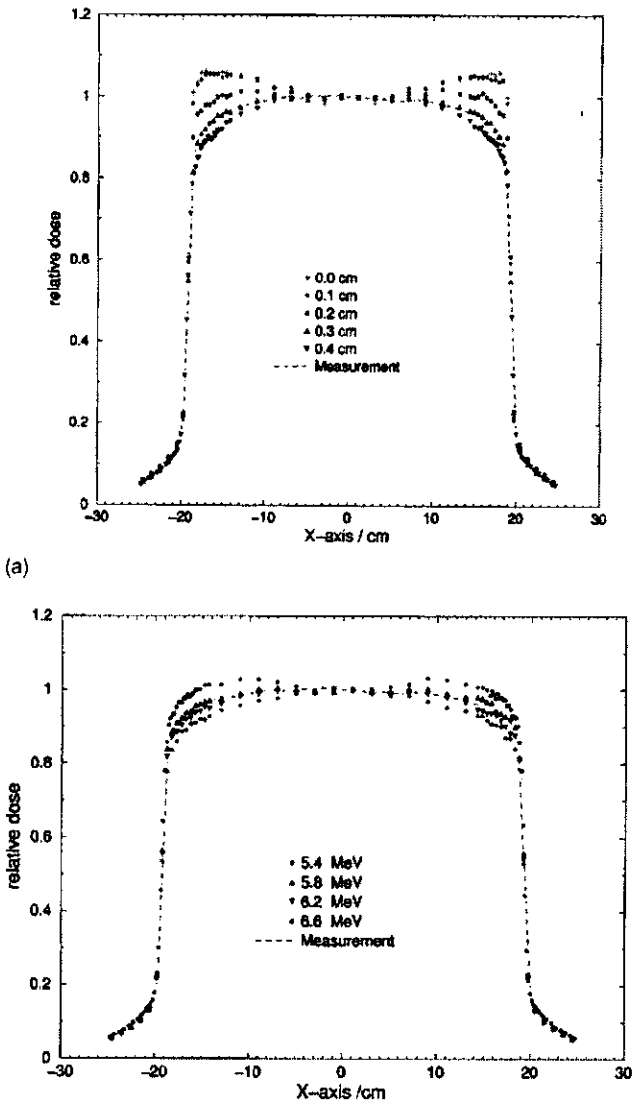


FIG. 3. (a) The lateral dose-profile curves as a function of the FWHM of the radial intensity distribution for a monoenergetic, 6 MV, electron beam. The measured profile and five calculated dose profile curves are presented; each curve is normalized to the central axis. (b) The influence of electron energy on dose-profile curves. The energy was varied from monoenergetic 5.4 to 6.6 MeV. For clarity only four calculated and measured profiles are presented; each curve is normalized to the central axis. The FWHM of the radial intensity distribution was 0.3 cm FWHM in all cases. Reprinted from Tzedakis *et al.* (Ref. 103) with permission.

the calculations.[102,104] The careful specification of the geometry and materials within the MC code is an important consideration. This includes verification of the input of the component geometric specifications, for example, by using the listing files available with the BEAM (Ref. 23) system. Moreover, for complicated components, such as the flattening filter, independent attenuation calculations of these structures, using monoenergetic photons, may be helpful to verify their thickness profiles.[105,106] Attenuation of the photon beam by the monitor ion chamber and the field mirror is negligible and these structures are often omitted from MC simulations of photon beams, except when backscatter to the monitor chamber is being taken into account. The ion chamber and mirror are, however, potentially important structures in electron beam simulations. The collimating jaws have no significant effect on the energy and angular photon distributions.[14,107,108] Changing the composition of the secondary collimators from pure tungsten to an alloy containing 50% tungsten by weight does not significantly affect the calculated dose.[102]

Even in those cases for which the beam model includes an explicit simulation of the accelerator head, end users of MC treatment planning algorithms will probably have little control over parameters affecting the treatment head simulation. It will be incumbent upon vendors to provide accurate beam models (verified by measurements) for treatment planning purposes. Vendors and developers should be aware of potential difficulties associated with accurate specification of the treatment head component structures. Such issues include difficulty in obtaining proprietary information from accelerator manufacturers, incorrect proprietary information provided by the manufacturers,[101,102,104] undocumented accelerator updates, and large uncertainties in important parameters needed for accurate simulation, such as the electron-on-target energy. Accelerator vendors are encouraged to make accurate, detailed information of their accelerators accessible in formats easy to implement in MC simulation, as established for instance by Siemens.[109]

III.B.2. Electron beam specifics

Simulation of the passage of electron beams through the treatment heads of accelerators has figured prominently in MC radiotherapy calculations for many years. The first application of the BEAM code[23] was to simulate electron beams for a wide range of accelerators. A variety of publications have demonstrated the accuracy of this technique for computing dose distributions and output ratios.[24,110,111] Recent work has established the methodology to achieve high accuracy in matching calculated and measured dose distributions for even the largest fields, including asymmetries and the bremsstrahlung tail.[112]

The procedure for simulating the electron beam treatment head is quite similar to that of x rays. The components generally important for electron beam therapy treatment head simulations are shown in Table III. Photon beam components are also presented for comparison. For electrons, accelerators from the major vendors use a pair of scattering foils to flatten the beam with minimal bremsstrahlung contamination. A low-scatter monitor chamber may be employed and the mirror may be retracted from the beam. The jaw position is generally fixed for a given applicator and energy such that the relative output ratio depends only on the custom insert.

The critical parameters for electron beam simulation are different than those for x rays. Sensitivity analyses provide quantitative evidence of these differences.[100,101,113] Due to the sensitivity of the beam range to the primary electron energy (a 0.2 MeV change in electron energy corresponds with a 1 mm change in beam range), the incident electron energy is the primary tuning parameter for electron beam

TABLE II. A summary of the findings of Sheikh-Bagheri and Rogers (Ref. 102).

Parameter in the linac model	Impact on in-air off-axis factors	Impact on central-axis depth dose values
Mean energy of the incident electron intensity distribution (assumed Gaussian)	**Decrease with increasing primary electron energy**, e.g., -0.105 ± 0.007/MeV at 15 cm off-axis for a Siemens KD 6 MV beam. **A 0.2 MeV mean energy change produces an observable effect.**	**A 0.2 MeV change in mean energy causes an observable change** (2% or 3σ with 0.7% or 1σ dose uncertainty).
Gaussian width of the incident electron energy distribution	**Show little or no dependence**, e.g., widening of the FWHM from 0% to 20% resulted in no change, for a Siemens KD 6 MV beam. Asymmetrical energy distribution has a small effect, e.g., an asymmetric Gaussian with 14% FWHM on the LHS of the peak and 3% FWHM on the RHS of the peak causes a change of 2% for a Siemens KD 18 MV beam.	**Show weak dependence in the dose buildup region and at large depths**, e.g., an asymmetric Gaussian with 14% FWHM on the LHS of the peak and 3% FWHM on the RHS of the peak increases buildup dose by up to 1.5% for a Siemens KD 18 MV beam.
Gaussian width of the incident electron radial intensity distribution	**Decrease quadratically with increasing Gaussian width**, e.g., a change in FWHM from 0.01 to 0.15 cm leads to 7% decrease at 15 cm off-axis for a Varian 18 MV beam.	**Little or no observable effect** considering statistical uncertainties.
Divergence of the incident electron beam (at a given intensity distribution FWHM)	**Show little or no effect up to 0.5°**; at 1° show a decrease of 1% at 15 cm off axis at 100 cm SSD, for an 18 MV Varian beam.	**No observable effect up to a few degrees** considering statistical uncertainties (1% or 1σ).
Radius of the upstream opening of the primary collimator	**Sensitive to small changes**, e.g., varying the upstream opening by 0.01 cm produces a 1% change at 15 cm off-axis for a Varian 18 MV beam.	No observable effect.
Density and material of the flattening filter	**Show strong dependence**, e.g., reducing tungsten density by 1 g cm^{-3} causes a 6% reduction at 15 cm off axis for a Varian 15 MV beam. **Using the incorrect material has a very large effect**, primarily because of the density change.	Not reported.

simulations. However, electron distributions are very sensitive to all the materials in the beam, especially the scattering foils and may also be affected by the monitor chamber if it has thick walls. Hence, accurate geometric descriptions of all components in the beam path are required for the simulation of electron beams. Asymmetries are evident in electron dose distributions and parameters to adjust include angle of the incident beam and the lateral position of shaped scattering foils and the monitor chamber.

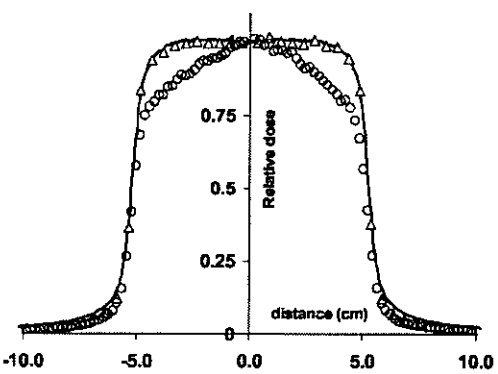


FIG. 4. MC calculated profiles in a water phantom (modified DPM, University of Michigan/UMPlan) for a 15 MV (10×10 cm^2) at 10 cm depth in water. Input for the profile calculations were the phase space simulations of the Varian 21-EX treatment head, performed using BEAMnrc, with two different material compositions specified for the flattening filter. Open circles represent the MC profile with a copper flattening filter (ρ = 8.933 g/cm^3) and open triangles that with the tungsten filter (ρ = 19.30 g/cm^3). Ion chamber measurements are shown in the solid line. Calculations and measurements are each normalized to the central axis (Ref. 104).

III.C. Modeling of the linear accelerator treatment head

III.C.1. General schemes

A beam model in the context of MC treatment planning is any algorithm that delivers the location, direction, and energy of particles to the patient dose-calculating algorithm. The direct MC simulation of a beam is one form of a beam model but for clarity we refer to it as a beam simulation rather than as a beam model. Accurate beam modeling is an important prerequisite for accurate dose calculation within the patient. Beam models use one of three possible ap-

TABLE III. Components in the treatment head and other beam modifying devices for x ray and electron beams.

X rays	Electrons
Exit window and target	Exit window and primary scattering foil
Primary collimator	Primary collimator
Flattening filter	Secondary scattering foil may be present
Monitor chamber	Monitor chamber (low scatter)
Mirror	Mirror
Asymmetric jaws and MLC	Jaw position fixed for each applicator
Wedge, blocks, graticule	Applicator with insert
Bolus	Bolus, shielding on or below patient surface

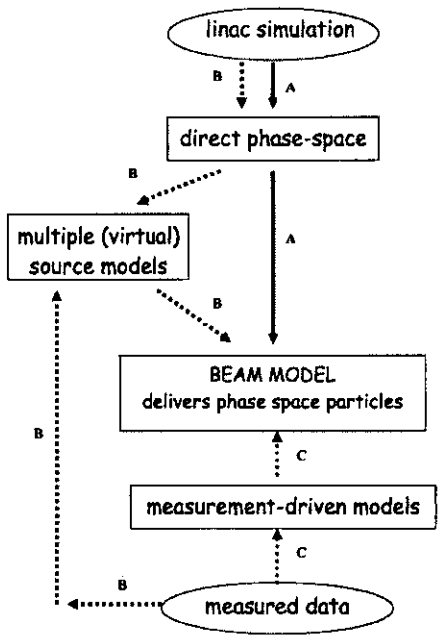


FIG. 5. The three "routes" for accelerator beam model specification: (a) solid line—direct use of phase-space information from simulation of the accelerator treatment head, (b) dashed line—multiple (virtual) source models derived from the phase-space information with or without enhancements from measured data, and (c) dotted line—development of other models derived from measurements (measurement-driven models).

proaches (see Fig. 5): (i) direct use of phase-space information from the accelerator treatment head simulation, (ii) development of virtual, multiple-source models reconstructed from the treatment head simulation with or without enhancement from measurements, or (iii) development of other models derived exclusively from measurements (measurement-driven models).

While the utilization of direct phase-space information provides details on the physical interactions within the treatment head, it may not be practical for routine clinical application.[65,93,114–116] Some of the limitations include: (a) generation and quality assurance of phase-space information requires MC simulation expertise, (b) accurate phase-space simulation is dependent upon accurate input parameters (such as the incident electron energy) as well as detailed geometric and material specifications of the accelerator components, which may be subject to inaccuracies, as well as proprietary issues, (c) storage requirements are large—typical MC simulation may require up to 10^9 phase-space particles for multiple photon beams,[115] which may amount to tens or hundreds of gigabytes of computer disk space for an accelerator with two photon energies and five electron energies with five different applicator sizes,[117] and (d) due to the much slower pace of speed increase in hard drives and networks compared to CPU speed development, reading the phase-space data can be a bottleneck in the calculation. It is therefore clear that a more concise, accurate characterization of radiation interactions in the treatment head is necessary for performing routine clinical MC dose calculation.

Another method for beam model specification, initially proposed by Ma *et al.*,[114] is based on the development of multiple-source models with parameters derived from the original simulated phase-space data.[65,88,93,114,116–120] In a multiple-source model, source particles are grouped by the location of their last interaction prior to being scored at the phase-space plane, resulting in subsources representing the major components of the treatment head. Fluence distributions for each subsource may be reconstructed from the phase-space data in the form of correlated histogram distributions,[65,93,114] thereby approximately retaining correlation of the particle's position, energy and direction. Although the development of accurate multiple-source models is also reliant upon simulation parameters and geometric and material specifications of the component structures, sub-sources may be "adjusted" (without redoing the simulations), based on measurements to optimize agreement between calculations and measured data, An example of such an approach consists of adjusting the spatial or energy boundaries of the planar fluence and energy distributions for each subsource to produce agreement with measured dose profiles and depth-dose curves, respectively, for a range of square field shapes, starting with fluence distributions reconstructed from phase-space data for one field size.[117,121] For instance, one changes the range of energies sampled from the discrete energy distribution as a function of radius, to provide a match with measurements, without having to redo the entire treatment head simulation. This method has been extended to MC-based commissioning of electron beams.[111,121] Models developed for reference accelerators were used to accurately commission (within 2%/2 mm of measurements in the high dose/high gradient profile regions) electron beams from other machines of similar design by adjusting a few parameters in the model. Implementation of photon and electron beam multiple-source models has been reported for a variety of accelerator manufacturers including, Varian,[122,123] Siemens,[73,124,125] Elekta,[73,86,88] Novalis (BrainLab),[87,89] and Cyberknife.[126]

A possibly more practical approach to beam modeling involves derivation of the model parameters from a standard set of measurements. The advantage of such measurement-driven models is that they may be developed without dependence on the details of the accelerator treatment head. Fluence distributions in measurement-driven models may be developed starting with analytical models whose parameters are optimized based on minimization of the differences between calculations and measurements.[100,125,127–131] Information for the model can be deconvolved from measured data.[100,127] Similar methods have been reported for beam-model specification using conventional dose calculation algorithms with good agreement established between calculations and measurements in water phantoms.[132–135] The same beam modeling criteria for conventional dose algorithms should also be applicable to the MC method although differences in the technical details such as source parameterization and software implementation are expected. The use of measurement-driven beam models eliminates the need for phase-space simulation and, as the model parameters are derived from measurements, may provide an added level of

confidence in the accuracy of the model in homogeneous phantoms. Moreover, specification of beam models is similar to that with conventional dose algorithms, and does not require expertise with MC accelerator simulation. However, rigorous verification of measurement-driven models in heterogeneous phantoms is necessary to validate the (empirically derived) fluence distributions. It has been shown that, in addition to the use of measurements in water phantoms, test cases in heterogeneous phantoms under conditions of electronic disequilibrium may be necessary to determine the correct energy spectrum unambiguously.[136]

III.C.2. Patient-specific beam modifiers

Transport through the patient-dependent components (such as the field-defining collimators and the MLC) may be classified into one of the following three schemes, which we will term: (a) explicit transport, (b) explicit-approximate transport, and (c) pseudo-explicit transport. In an explicit transport scheme, all particles (with appropriate energy cutoff values) are transported using MC techniques through the components; all details of the design geometry (such as rounded leaf ends, interleaf spacing, etc. for a MLC) should be included in the geometry modeling.[72,115,137] With explicit-approximate transport, approximations are employed in the MC photon/electron tracking scheme to improve the efficiency of the calculation.[83,87,138,139] An example is the approach of Siebers *et al.*,[83] in which only first Compton scattered photons are transported through the MLC. The method of Tyagi *et al.*[139] includes simulation through the detailed MLC geometry accounting for all Compton scattered photons, however, ignores the secondary electrons, i.e., assumes they deposit their energies locally. In a pseudo-explicit transport scheme, beam fluence distributions are reconstructed from the phase-space simulation to develop subsources for characterizing components, such as the field defining jaws,[93,120] electron applicators,[114] and the MLC.[88] Although the need for time-intensive, explicit transport is obviated with the use of multiple subsources, the ability to incorporate detailed geometric characteristics of components like the MLC may be difficult with this approach. Appropriate benchmarking must be performed by developers and vendors to evaluate the tradeoffs between speed and accuracy related to the use of various beam model approaches in MC-based modeling of patient-specific beam modifiers.

III.C.3. Output ratios

The output ratio (or relative output factor) is the ratio of dose per monitor unit in a phantom for an arbitrary field size to that for a reference collimator setting, the latter usually 10×10 cm^2 at a SAD (or SSD) of 100 cm for a depth of 10 cm for x-ray beams.[140] Output ratios over the range of clinically useful field sizes are heavily influenced by the amount of head scattered radiation impinging on the phantom.[140,141] Radiation generated by clinical accelerators can be characterized by a primary photon source generated through the bremsstrahlung process, and other extrafocal sources accounting for scattered photons arising primarily from the flattening filter and primary collimator, typically 3%–8% of the energy fluence in a 10×10 cm^2 field. The primary source is a narrow sharp source normally a few millimeters in diameter, while the extrafocal sources tend to be much broader, more diffuse and bell-shaped in nature.[142,143] The output ratio is affected by the size and geometry of these sources. The collimators (jaws and MLCs) block the extrafocal source when the field size is sufficiently small ($<3\times3$ cm^2) and for smaller fields the primary source starts being blocked by the collimators. The combination of extrafocal source and primary source cutoffs cause the output ratios to fall off sharply at small field sizes.[143,144] Accurate source modeling of small fields is especially important in IMRT planning where multiple, small field, off-axis segments are often used.[61,87,145] For medium sized fields (10×10 cm^2–20×20 cm^2), output ratios are more likely affected by the extrafocal sources as the collimators are large enough to expose the entire primary source, however, small enough to eclipse the extrafocal sources. In large fields, beyond 20×20 cm^2, the extrafocal sources are nearly completely exposed—the increase in output ratios in these situations is primarily a result of phantom scatter, lack of backscatter from the collimator jaws to the monitor chamber, and stray radiation from the treatment head.[146,147] Depending on the location of the jaws relative to the transmission chamber (which varies among the different linear accelerator types), radiation backscattered from the jaws into the chamber can also have an effect on the output ratios. Relative to large field sizes (40×40 cm^2), backscattered radiation increases by 2%–3% at small field sizes (3×3 cm^2) for 15 MV photons on a Varian (21EX) linear accelerator, for example. Studies for photon beam MC algorithms have shown calculated output ratios to be within 1.5% of measurements over a range of field sizes when accounting for backscattered radiation.[27,139,148,149] MC calculated electron beam output ratios (for field size-specific applicators) have also been reported to agree with measurements within 1%–2%.[26,110]

III.C.4. Dose buildup

There has been considerable discussion on discrepancies between MC calculations and measurements in the dose buildup region. Hartmann-Siantar *et al.*[27] observed MC dose deficits of up to 5 mm distance-to-agreement versus measurements and attributed these discrepancies to a source of electrons in the accelerator head not fully accounted for in the treatment head simulation with BEAM. It has been shown that arbitrarily increasing the electron contamination by a significant amount removes the discrepancy.[27,88,125] As the result of further investigation, Ding[150] concluded that the discrepancy could not be explained by the electron source hypothesis and postulated that contaminant neutrons emerging from the treatment head might be the cause. However, Ding *et al.*[151] subsequently refuted this neutron hypothesis by measuring minimal neutron component in the dose buildup region. Abdel-Rahman *et al.*,[152] who performed MC calculations (using EGSnrc) and a comprehensive set of measurements with multiple detectors, again found significant differ-

ences between measurements and calculations in the buildup region for 18 MV photons even when completely simulating the response of the detectors. They ruled out the following possible causes of the discrepancy: (a) an unknown electron source in the accelerator head simulation, (b) contaminating neutrons, (c) inaccurate cross section data, and (d) (gamma, p) reactions. They also showed that explicit modeling of triplet production interactions influences the dose buildup for an 18 MV beam. However, work by Kawrakow[153] showed that a more accurate triplet production model does not remove the discrepancies. Benchmarking of the NRCC accelerator photon beams has shown agreement well within 1%, for field sizes up to 10×10 cm^2, at all depths[154]; it should be noted that the NRCC accelerator (20 MV photons) uses a sweeping beam technique[154] as opposed to a flattening filter to flatten the beam. These comparisons explicitly accounted for stopping-power ratio variation with depth.

More recently, Kawrakow,[155] in performing detailed ion chamber simulations using the EGSnrc code system, showed that the relationship between measured ionization and dose for relative photon beam dosimetry depends on details of the chamber design, including cavity length, mass density of the wall material, size of the central electrode, and cavity radius, in addition to the beam quality and field size. When the correct ionization-to-dose relationship was used with the experimental data[155] and a variety of other improvements in the head simulations were made (e.g., using a larger diameter primary collimator opening and including the effects of extra shielding upstream of the jaws or MLC,[156] correcting a bug in the JAWS component of the BEAMnrc code, including an angular spread in the incident electron beam and several small effects in the simulation[153]), the discrepancies in the build-up region were reduced to an acceptable level. Chibani and Ma showed that resolving inaccuracies in the modeling of the primary collimator for a Varian 18 MV accelerator as well as including virtual sources for the lead shield and mirror frame, resulted in significantly better agreement between calculations and measurements in the dose buildup region.[156] Other sources of inaccuracy associated with detectors for dose buildup measurements are presented in Sec. III E 3.3.

III.D. Treatment planning: MC-based patient calculations

III.D.1. Statistical uncertainties

III.D.1.a. Latent variance and statistical estimators. For a finite number of independent simulated histories (N), the dose calculated using the MC method is subject to statistical uncertainty. By invoking the central limit theorem,[157] one can show that the statistical uncertainty in dose is proportional to $1/\sqrt{N}$, in the limit of infinite (large) N. There are generally two sources of statistical uncertainty in MC calculations of patient dose—those resulting from the simulation of the accelerator treatment head and those arising from fluctuations in the phantom/patient dose calculation. Sempau *et al.*[95] coined the term, "latent variance" to describe the uncertainty due to statistical fluctuations in the phase-space as opposed to the uncertainty due to the random nature of dose deposition in the phantom. Using an already calculated phase-space file, the statistical uncertainty in the dose calculated in a phantom by reusing the particles from the phase-space file (i.e., assuming they are independent and ignoring correlations between them), will approach the finite, latent variance associated with the phase space data, regardless of the number of times the phase space is reused. The use of source models derived from phase space simulation will tend to smooth out point fluctuations in the phase space.[114] However, if the latent variance is large enough to introduce systematic bias then this will be propagated in the reconstructed phase space (source model). Beam models derived exclusively from measurements, on the other hand, are analogous to those generated using conventional (analytical) algorithms—latent variance (as defined above) is not a concern for such models, but other systematic uncertainties in the beam model will be present. In estimating the statistical uncertainty in the patient dose calculation, it is necessary to account for the latent variance from the phase-space calculation as well as the random uncertainty from the patient calculation. To make this possible in practice, more work is needed to develop tools to assess the role of latent variance in patient dose calculations. Should latent variance be a significant factor in the total uncertainty, more independent phase-space particles need to be used in the patient simulations. It must be emphasized that all beam models are subject to systematic uncertainties, which are analogous to those introduced by the latent variance. For measurement-driven models, these uncertainties will be related to inaccuracies in the measurement data. Both types of models are subject to systematic uncertainties due to inadequacies in the model itself.

There are two common methods for calculating statistical uncertainties: the batch method and the history-by-history method. In the batch method, the estimate of uncertainty (standard error of the mean, $s_{\bar{x}}$) of a scored quantity, X, is given by

$$s_{\bar{x}} = \sqrt{\frac{\sum_{i=1}^{n}(X_i-\bar{X})^2}{n(n-1)}}, \tag{3a}$$

where n is the number of independent batches, X_i is the scored quantity (such as dose) in batch i, and $\bar{X}$ is the mean value of X over all the batches. The sample size is therefore given by the number of batches, where each batch is a calculation of the same quantity carried out with independent phase-space file inputs and random number sequences. In the history-by-history method, X_i represents the scored quantity in history i (rather than batch i) so that the standard error of the mean can be recast (in a mathematically equivalent form) as follows:

$$s_{\bar{x}} = \sqrt{\frac{1}{N-1}\left(\frac{\sum_{i=1}^{N}X_i^2}{N}-\left(\frac{\sum_{i=1}^{N}X_i}{N}\right)^2\right)}, \tag{3b}$$

where N is the number of primary (independent) histories, X_i the contribution to the scored quantity by independent his-

tory, i. An issue evident with the batch approach [Eq. (3a)] is that the sample size, n, is given by the number of batches. As n is usually small (on the order of ten or less) there is statistical fluctuation in the uncertainty itself. Advantages of the history-by-history method have been detailed elsewhere.[158]

An important consideration when calculating uncertainties is to take into account the correlation between a primary particle and all its secondaries, especially in the case of bremsstrahlung splitting where a large number of photons may all come from a single electron. Thus, to be strictly correct, these secondaries must be treated as part of the same history. If this correlation is not taken into account one can underestimate the uncertainty in a dose calculation, as the secondaries will be treated as independent particles thereby reducing the uncertainty erroneously. Another important case of correlation is when a single particle is being used several times as a source particle. In this situation it is important to recycle the particles, i.e., use them multiple times, all at once and treat them as part of the same history. If one were to restart the phase-space file multiple times, one would lose the correlation between particles which are all part of the same history. These secondaries would then be treated as independent particles causing the uncertainty to be underestimated. A more comprehensive review of the implications of recycling and restarting phase-space particles is provided in the paper by Walters *et al.*[158]

III.D.1.b. Influence of statistical uncertainties on dose distributions. For radiation therapy dose distributions, $s_{D_i} \propto \sqrt{D_i}$, where s_{D_i} is an estimate of the standard error of the mean (standard deviation/$\sqrt{N}$) of the dose in voxel i and D_i is the dose in that voxel.[159–161] The fractional (or relative) uncertainty in dose, $F_{D_i}=s_{D_i}/D_i \propto 1/\sqrt{D_i}$. In other words, the fractional uncertainty in dose in a voxel decreases as the dose increases. This relationship provides a useful rule of thumb when viewing dose distributions since it implies that the relative uncertainty in the dose in high-dose regions will be smaller than in low-dose regions, even though the absolute uncertainty is usually larger.

From Eq. (3b) we see that the uncertainty is roughly proportional to $1/\sqrt{N}$. Since, the simulation time $T \propto N$, it can be seen that achieving absolute precision ($s_D=0$) with MC simulation requires an infinite calculation time. Fortunately, absolute precision is not required in dose calculation results. This section concerns the precision required for MC simulation and the impact of a lack of precision on MC dose distribution quantities.

Radiation therapy dose distributions contain many voxels in which the dose is computed (for example, a cubic volume with sides of 10 cm contains 15 625 voxels with sides of 0.4 cm). Since the subvolume receiving a therapeutic dose can be highly variable, a standardized method to specify the statistical uncertainty for such a distribution is necessary. Although the statistical uncertainty can be specified for a single voxel in a dose distribution (such as that at the isocenter or at the maximum dose voxel, $D_{\max}$), as described below, these voxels are poor measures for the uncertainty of a MC computed plan. Alternatively, the statistical uncertainty over some volume, such as a planning target volume or some dose volume, such as the volume receiving greater than $X\%$ of the treatment dose, can be computed from the square root of the average variance of each constituent voxel. For example, the fractional uncertainty in the average dose for voxels with dose values greater than 50% of the maximum dose, $\bar{F}_{D>0.5D_{\max}}$ as suggested by Kawrakow, and Rogers and Mohan[99] could be used

$$\bar{F}_{D>0.5D_{\max}} = \sqrt{\frac{1}{K_{D>0.5D_{\max}}} \sum_{D>0.5D_{\max}} \left(\frac{\overline{s_{D_i}}}{D_i}\right)^2}, \tag{4}$$

where D_i is the dose estimate in the ith voxel, $\overline{s_{D_i}}$ is its uncertainty, and the summation runs over the K voxels with dose greater than 50% of $D_{\max}$. For clarity in communication, this report recommends that quantities, such as $\bar{F}_{D>0.5D_{\max}}$ or $\bar{F}_{\mathrm{PTV}}$ (or $\bar{F}_{\mathrm{PRV}}$) for doses to the specific volumes, planning target volume (PTV) or planning risk volume (PRV), be adopted as a standard method of reporting statistical uncertainties in dose averaged over the relevant volume. The use of uncertainties to voxels, such as the maximum dose should be avoided. In situations where doses in single voxels are important, such as the maximum dose to a "serial" organ like the spinal cord, users are reminded to also consider the statistical uncertainty of that dose voxel.[162] In such instances, it may be necessary to simulate a large enough number of histories so that s_{D_i} is very small.[162] This will ensure that the absolute uncertainty in the highest dose voxel will also be small. The statistical precision required for dose estimates in single voxels should be decided upon with guidance from the clinician. Users should also note that for a constant number of source particles, the statistical uncertainty also depends upon the size of the dose voxel. Reducing the volume to achieve a "point-like" voxel will require increasing the number of particles simulated to achieve constant statistical precision.

To varying degrees, statistical uncertainties affect all measures of the dose distribution, including output ratios, isodose profiles, dose-volume histograms, dose response parameters such as equivalent uniform dose and TCP/NTCP. In a uniform dose distribution, the dose metrics most sensitive to statistical uncertainties are the maximum and minimum dose voxels. These extreme values are by definition the voxels that have the greatest deviation from the mean dose. If one desires a region of uniform dose (e.g., within the PTV), that dose will not be uniform using a MC-based calculation, due to statistical fluctuations between adjacent dose voxels. The reported minimum and maximum dose voxels will differ from the mean of the idealized dose distribution by up to several standard deviations if many voxels are present in the distribution.[45,163] For example, in a uniform dose distribution with 15 625 voxels computed with MC algorithms, due to statistical fluctuations, there is a 63% probability that the dose in at least one voxel differs from the mean by more than four standard deviations.[45] With regard to dose prescriptions, the specification of the maximum dose to a single voxel when there is a desired volume of uniform dose (e.g., the

PTV), will result in an underdosage to this volume. Similarly, dose prescription to the minimum dose voxel will result in an overdose to the relevant volume. Isodose contours are also sensitive to statistical noise. For well-defined fields, such as rectangular fields used in beam commissioning, even 1% (1s) statistical uncertainty causes observable jitter in isodose contours. For patient fields that have irregular shapes, the acceptable amount of statistical uncertainty for isodose viewing is a matter of personal preference and should be agreed upon by the planner and the physician. It has been suggested that 2% statistical uncertainty on the D_{max} voxel is adequate for isodose evaluation.[160] When viewing statistical jitter in MC isodose distributions, the physicist should remind observers that overall dose delivery accuracy is limited to within a few percent; therefore there is uncertainty in the actual location of an isodose surface even when dose is computed with non-MC algorithms. From this point of view, the planning team can use MC isodose jitter as a mechanism to open the dialog on realistic dose uncertainty in actual treatment delivery.

Integrated dose quantities, such as dose volume histograms (DVHs) are less sensitive to statistical uncertainty.[159–165] DVHs computed with the MC method represent the actual DVH (that computed with a hypothetical 0% statistical uncertainty simulation), convolved with a statistical uncertainty distribution. With this realization, Sempau and Bielajew[159] and Jiang *et al.*[165] reported that the effect of the "statistical noise" on DVHs could be removed by deconvolving the uncertainty from the DVH, allowing substantial decrease in the required number of histories used, depending on the complexity of the DVH. In general, the blurring effect due to statistical noise is greatest for steep DVHs, such as those for PTVs, while shallow DVHs, such as those for critical structures are less affected.

The sensitivity of quantities such as TCP and NTCP to statistical noise depends upon the parameters used by the model[164] and on the magnitude of the noise. Kawrakow[161] has shown for general dose-based cost functions that the uncertainty in the cost function decreases more rapidly than the individual dose uncertainties when the plan is close to optimum as expressed by the cost function. This implies that during IMRT optimization, where plan updates are based upon evaluation of such cost functions, larger statistical uncertainty might be acceptable.

A method to reduce the effect of statistical uncertainties in MC dose distributions is to postprocess the dose distribution. These methods have been termed denoising or smoothing techniques. Various methods related to digital filtering,[166,167] wavelet thresholding,[168,169] adaptive anisotropic diffusion,[170] and denoising based on a minimization problem[171] have been proposed. A detailed comparison of these methods is presented in the article by El Naqa *et al.*[172] Denoising is an approximate, efficiency enhancing method (i.e., it is not a variance reduction technique) since it can introduce systematic bias into the calculation. Nonetheless, denoising techniques are useful as they can reduce the overall (systematic +random) uncertainty when the random component decreases more than the systematic component increases. Kawrakow presented a series of tests to evaluate the suitability of MC dose distribution denoising algorithms.[167] These tests are based on the fact that one can determine the "correct" dose distribution by simulating many histories (for a very high precision) and then comparing the denoised distribution from a simulation with a much smaller number of histories to the high precision results. The tests include:

- visual inspection of isodose contours,
- evaluating root-mean-square difference between dose distributions,
- evaluating the maximum dose difference,
- comparisons of dose-volume histograms with and without denoising, and
- comparing the fraction of voxels failing an $x\%/y$ mm test.

Denoising methods reduce the number of particles (and, hence, the calculation time) required to achieve a given uncertainty by a factor of 3–10. Denoising techniques require proper validation under the full range of clinical circumstances before they are used with MC dose algorithms.

III.D.2. Dose prescriptions and monitor unit calculation

The stochastic nature of the MC method raises questions for prescribing dose. It is common clinical practice to prescribe dose to a single voxel or to base the dose prescription on the maximum or minimum dose voxels. However, as discussed above (Sec. III D 1.2), in an approximately uniform dose distribution, the outliers (the maximum and minimum dose voxels) are subject to the largest statistical fluctuation and even other single voxel doses may lack the precision for monitor unit calculations. Standard treatment planning analysis methods using isodose distributions and dose-volume histograms rely on dose averaged over a volume. It is logical to extend this practice to the prescription of dose, thereby avoiding precision issues in doses calculated in small volumes. For example, dose may be prescribed to an isodose surface, to a region of uniform dose (averaged over many voxels) about the isocenter, or to a single point on a dose volume histogram. The practice is emerging to calibrate the calculated dose distribution by performing a calculation in the standard geometry where the accelerator is set to deliver a given dose per monitor unit, (e.g., 1 cGy/MU) at a given voxel. In this case, the dose at the calibration point may be calculated to high precision by averaging over a large number of voxels in a uniform dose region.

The issue of monitor unit calculations to specific single voxels within the target volume (e.g., the isocenter) for routine clinical planning may be confounded by large statistical fluctuations in the doses to individual voxels. Although the treatment planning practice is quickly moving toward volume-based dose prescriptions, particularly for IMRT planning, the ability to perform second MU checks for plan verification purposes is an important component of the standard practice. Until more efficient solutions are available, MU cal-

culations based on single-voxel-dose prescriptions should be performed with very high precision. That said, this task group encourages a shift in paradigm from point-based toward volume-based dose prescriptions, which we feel will soon become the standard method for prescribing doses in radiotherapy planning. The task group strongly discourages vendors from using the D_{max} or D_{min} dose voxels or other single voxel doses for dose prescription and monitor unit calculations in their MC-based treatment planning systems. Current users of MC algorithms are encouraged to find ways of circumventing point-based dose prescriptions if their systems are not flexible enough to allow otherwise. To obviate the concerns of dose prescription based on a single voxel, one institution (author J.E.C., Ottawa Hospital) has developed the following "work around" method for MC-based electron beam calculations: A dose distribution is calculated using 100 MU; this dose distribution in absolute terms is equivalent to a relative isodose distribution normalized to the standard calibration conditions (10×10 applicator, 100 cm SSD, d_{max} depth). For dose prescription, the physician chooses the isodose line that encompasses the target. The dose prescription point is then positioned on this isodose line along the central axis of the beam, and the treatment MUs are determined. A second calculation is performed as a check of the MUs and the final dose distribution.

For multiple 3D-CRT or IMRT fields, the procedure for generating monitor units is similar to that for single fields. For example, a given isodose line (such as the 95% line) can be selected for dose prescription and, for a given field, the dose contribution to a single voxel (e.g., the projected field cax point) along this line is determined from the treatment plan. For a calibrated MC algorithm, the absolute dose contribution from the given field in this voxel (on the selected isodose line) will be computed in units of cGy/MU, from which the monitor units for the beam can be calculated. A complete formalism for MU calculations for the different types of treatment deliveries has been provided by Ma *et al.*[173]

III.D.3. CT-to-material conversions

For conventional algorithms, electron densities extracted from the CT image are used to scale the influence of primary and, ideally, also secondary radiation interactions. MC algorithms utilize material density and the material atomic composition when performing particle transport. The differing atomic compositions of patient materials (e.g., soft tissue, bone, lung, air) result in different cross sections for the various radiation interactions. While material compositions cannot be determined solely from a single energy CT, they can be indirectly approximated by estimating the mass density from the electron density followed by assigning a material to each voxel.[37,86,174–176] For some MC codes, explicit material specification is circumvented by directly relating the CT [Hounsfield (HU)] numbers to material interaction coefficients, based upon parameterization of materials representative of the patient,[37] for example those tabulated in ICRU Report No. 46.[177]

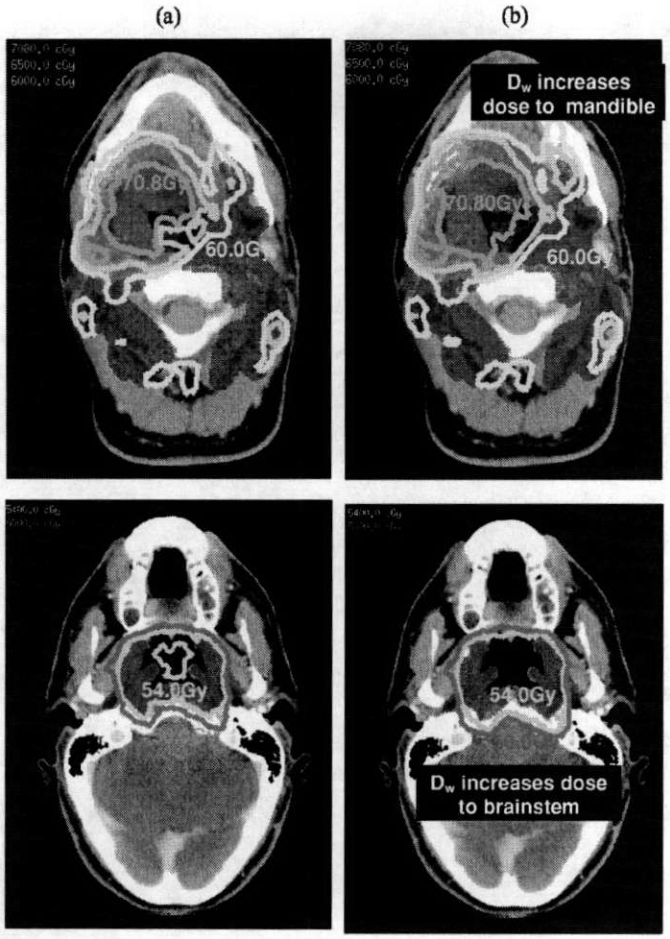
To ensure proper correspondence between HU and materials (or material interaction coefficients), correspondence between these quantities must be established during the CT and treatment planning system commissioning process. To ensure appropriate material specifications, it may be desirable to have several conversion tables, whose selection is based upon knowledge of the particular patient's characteristics. Use of multiple calibration curves reduces the volume of inappropriate tissues specified in given regions, such as lung tissue within a prostate gland, however. Although the importance of exact material specification has been established in other studies,[178] more work in this area of research relevant to clinical treatment planning is necessary. Other information about the patient may also be helpful. For example, in a patient with a hip implant, it may be impossible to distinguish if the implant is titanium, steel, or a composite based solely on the CT numbers. In this case, material assignment based on knowledge of the material implanted in the patient would be beneficial. A recent evaluation of the influence of material compositions on dose distributions[178] showed dose errors of up to 10% for 6 and 15 MV photons, and 30% for 18 MeV electrons due to media and/or mass density misassignment, when comparing dose distributions between a known phantom and a CT-imaged phantom with compositions and densities assigned by a conversion process. The use of conversion techniques based purely on mass density (e.g., assuming the only patient material is water, but with varying density), as employed in conventional algorithms, is discouraged with MC simulation because most of these methods ignore dependencies of particle interactions on the materials, which can lead to notable discrepancies in high atomic number materials.[176]

CT number artifacts caused by issues such as beam hardening in the CT scanning process or by high density structures, such as dental fillings are potentially important in MC dose calculation. Other artifacts may arise, for example, when the CT scanner encounters a sharp edge, such as the surface of a rectangular solid phantom, where a blurred edge may result after image reconstruction. In one observed case, the resultant contour showed 3 mm of additional phantom, causing the MC-calculated percentage depth dose curves to be shifted 3 mm toward the surface.[269] Such issues are of relevance to both MC- and non-MC-based algorithms and will need to be taken into consideration in order to perform accurate dose calculations in the dose buildup region.

As with any dose algorithm, testing should be performed to evaluate the effect of artifacts on the accuracy of the MC dose calculation.[179,180] It must be emphasized that the accuracy of CT-number to material conversions affects all dose calculation algorithms, both MC- and non-MC-based methods.

III.D.4. Dose-to-water and dose-to-medium

Historically, radiotherapy dose measurements and calculations have been performed in, or specified in terms of the absorbed dose to water (D_w). With MC-based algorithms,



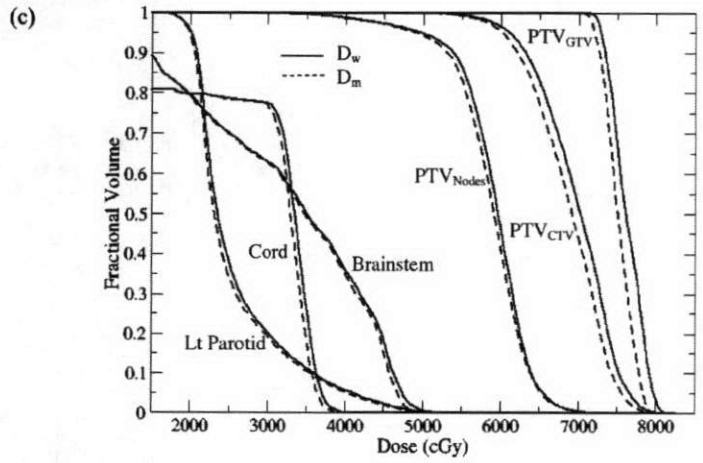


FIG. 6. MC-based (MCV system) dose-to-medium (D_m) and dose-to-water (D_w) results for a typical head and neck IMRT treatment plan: (a) D_m, (b) D_w, and (c) DVH comparison. Reprinted from Dogan *et al.* (Ref. 181) with permission.

particle transport simulations occur in materials representative of patient media; dose is therefore specified to the patient medium (D_m). For tissues with densities near 1.0 g/cm^3, the difference between D_w and D_m for megavoltage photon beams is small (1%–2%), however, for higher density materials, such as cortical bone, the difference can be as large as 15%,[176] since the stopping powers of water and these higher-density materials differ more significantly. Therefore, there is a systematic difference between the dose computed using conventional analytical algorithms and MC simulation. Figure 6 shows the dose and DVH differences between plans calculated with D_w and D_m for a typical head and neck IMRT treatment plan.[181]

To use MC simulation in the current clinical practice so as to be able to compare D_m with historical D_w results, requires a conversion of D_m to D_w for dose prescriptions, isodose coverage, dose-volume histograms, and any other dose related metrics. In this context, the converted D_w represents the dose to a small volume of water embedded in the actual medium. Whether one should eventually use D_m in place of

D_w directly in clinical prescriptions remains the subject of debate.[182] Arguments in favor of using D_w for dose specifications include:

- Historical clinical experience has been derived based on D_w, hence, D_w allows direct compliance with previous clinical experience and with conventional dose algorithms. Doses reported in clinical trials are based on D_w, hence, therapeutic and normal tissue tolerance doses are based on D_w.
- Accelerator and ionization chamber calibration protocols are D_w based.
- Tumor cells embedded within a medium are more water-like than medium-like, e.g., a tumor cell embedded in a bone matrix.

Arguments in favor of using D_m for dose specification include:

- D_m (or the dose to the tissues of interest) is the quantity inherently computed by MC dose algorithms. This may be of more clinical relevance than the doses on which historical clinical experience is based, which are approximate estimates of the true dose in the first place.
- Converting D_m back to D_w may involve additional complexity and introduce additional dose uncertainty.
- The difference between D_m and D_w for tissue equivalent materials is rather small and is likely to have minimal impact in clinical practice.
- It is known that, due to organ and target motion, the dose actually delivered in the course of a treatment may be significantly different from the planned dose. Conversion of Monte Carlo computed doses to D_w may analogously necessitate discussions on conversion to "dose to a static patient," should 4D planning and delivery techniques become routinely used in the clinic.[183]

The conversion of D_m to D_w or vice versa requires an application of Bragg–Gray cavity theory: $D_w = D_m(\bar{S}/\rho)_m^w$, where $(\bar{S}/\rho)_m^w$ is the unrestricted water-to-medium mass collision stopping power averaged over the energy spectra of primary electrons at the point of interest. One method to accomplish this conversion[176] is to utilize the fact that for patient-like materials, $(\bar{S}/\rho)_m^w$ is approximately invariant throughout a photon radiation therapy field (within 1%), hence, $(\bar{S}/\rho)_m^w$ can be used as a post-processing step to convert D_m to D_w. However, this requires that a sufficient number of materials be specified so that the material-dependent dose conversion approximates a continuous function.[184] Rather than postprocessing the dose conversion, one can multiply the energy deposited by primary and secondary electrons on each electron energy-loss step by the factor $(L/\rho)_m^w$, the ratio of the restricted mass collision stopping powers of water to local medium for the current energy of the electron. This is done in the MC transport code, thereby directly obtaining D_w. Alternatively, material interaction data may be evaluated over continuously parameterized space with cross sections and stopping powers for specific materials scaled relative to water.[35,37] Figure 7 shows a parameterization of the mass collision stopping power as a function of tissue density for different electron beam energies.[37] The fit is shown to be in good agreement with ICRU data[177] for body tissues. The advantage of this approach is that it circumvents the potential misassignment of media at material boundaries[37] which arises when the human tissue mass density range is divided into bins representing different tissues, especially when the number of material bins is small.[184]

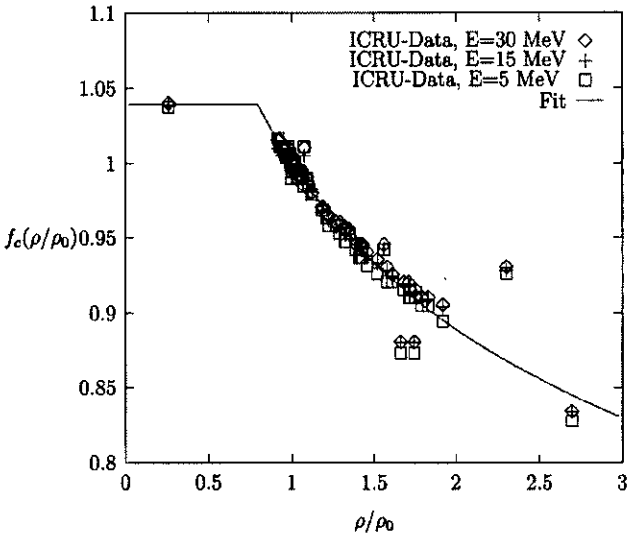


FIG. 7. Mass collision stopping power divided by the mass collision stopping power of water as a function of density normalized to water. A comparison of the fit function to the ICRU data (Ref. 177) for body tissues for electron energies of 5, 15, and 30 MeV is shown. The few points outside the curve are materials like urinary stones, which are considered to have negligible effects. Reprinted from Kawrakow *et al.* (Ref. 37) with permission.

Until further studies indicate clinical justification for selecting D_m or D_w, it is the consensus of this task group that MC dose results should: (a) explicitly indicate the material to which the dose is computed and (b) allow conversion between D_m and D_w using one of the methods discussed above, or other methods as developed in future investigations.

III.D.5. IMRT dose calculation and optimization

A strength of the MC method is its ability to accurately compute dose for complex dose delivery scenarios such as encountered in IMRT. IMRT often involves large intensity gradients and is usually delivered using a sequence of small static or dynamically shaped MLC segments. Under these circumstances, the assumptions used in conventional algorithms regarding scatter equilibrium and output ratio variation with field size often break down.[185] Additionally, in IMRT a significant fraction of dose to structures of interest (particularly dose limiting critical structures) is due to radiation scattered from or transmitted through the MLC.[186] MC simulation circumvents these limitations since it makes no assumptions regarding radiation equilibrium and can transport particles through the detailed MLC leaf geometry[87,137,187] to include MLC leakage radiation, even for dynamic MLC leaf sequences.[138,188,189]

In using MC calculation in IMRT, a consideration is the method used to incorporate the MLC into the dose calculation. Intensity modulation has been incorporated into MC simulation using the conventional planning system's intensity matrix,[190] independently generated fluence modification matrices,[87,92,191,192] or direct transport through the MLC.[61,189] Note that errors introduced by fluence approximations used during the MC dose calculation will be propagated through to the prediction of the patient's dose. Thus, when a fluence matrix approach is used for MC dose calculations, differences with respect to a conventional algorithm's heterogeneity correction will be detected, but fluence prediction errors may go undetected, particularly if the same fluence matrix is used both for the conventional and MC calculations. However, when MC simulation is used to transport directly through the detailed MLC geometry, these fluence errors should be detected. Accounting for geometric details in the MLC geometry, such as interleaf leakage is possible using a fluence matrix approach, however, will require a very high resolution calculation matrix, which may be a limiting factor. Moreover, the energies of the scattered particles through the MLC using such an approach will be approximate. Whether or not modeling the intricate details of the MLC (versus more approximate fluence matrix approaches) will lead to clinically significant differences in IMRT treatment planning is not a fully resolved issue. More treatment planning studies in this area of research are necessary to better understand the associated clinical implications.

Application of the MC method for IMRT QA has been demonstrated by several research groups[61,87,139,190,192–194] using MC simulation to recompute dose distributions optimized with the conventional dose planning algorithm. When using MC calculations as the reference plan and ignoring statistical fluctuations, dose differences between conventional and MC algorithms can be considered systematic.[195] Ma *et al.*[191] found dose errors in excess of 5% and 20% (relative to the prescribed dose) in targets and critical structures respectively due to patient heterogeneities, in comparing MC calculations (employing an independent fluence matrix) with a pencil-beam model. In comparing MC calculations with a conventional planning system's (pencil-beam model) intensity matrix for head and neck and lung cases, Wang *et al.*[190] found a 20% lower V_{95} (volume receiving at least 95% of the target dose) for a lung plan, and a 9% lower D_{95} (dose delivered to at least 95% of the target volume) for a head and neck plan. Average agreement among all head/neck and lung cases in this study, however, was quite good.[190] Regarding the transport of particles through the MLC, Siebers and Mohan[61] showed fluence-based IMRT dose underestimates of 4.5% in V_{95} and heterogeneity-based dose overestimates of 5% in V_{15} in the same treatment plan. In intercomparing MC codes, Reynaert *et al.*[196] reported deviations of up to 10% in DVHs between PEREGRINE and BEAMnrc calculations. These discrepancies were attributed to differences in modeling of the Elekta (SLi-plus) MLC, not caused by the particle transport in the patient. This study further illustrates the importance of careful modeling of the details of the MLC in IMRT treatment planning.

MC-based IMRT plan optimization is limited by the clinical availability of MC calculation as a whole and the large calculation time required to perform the multiple IMRT dose calculations required for optimization. MC simulation during optimization allows the optimizer to account for heterogeneity induced dose perturbations, as well as for MLC leakage and scattered radiation. Inaccurate dose algorithms used during optimization can result in convergence errors[195] in which the optimized fluence pattern differs from that corresponding to the optimal dose distribution.

Studies demonstrating the use of the MC method in IMRT optimization include the work of Laub *et al.*,[197] who utilized a MC algorithm to evaluate the cost function during optimization but a pencil beam to compute the cost function derivatives used by the optimizer. Jeraj *et al.*,[195] reported on convergence and systematic errors in the IMRT inverse planning process (for lung cancer) resulting from the use of convolution/superposition and pencil beam algorithms, versus the Monte Carlo method. A similar study for treatment planning of head and neck cancers was published by Dogan *et al.*[198] Another study by Siebers *et al.*[199] demonstrated the use of correction-based schemes to produce convergence of doses computed by conventional algorithms with MC calculations. Bergman *et al.*[200] reported on the use of an EGSnrc-based MC beamlet dose distribution matrix for IMRT planning using a direct aperture optimization algorithm. The goal of their work was to assess the improvement in accuracy over conventional algorithms in using MC methods for both the final dose calculation as well as in the inverse planning process.[200] Combining these methods with statistical smoothing and denoising techniques,[166–172] after comprehensive benchmarking, may allow introduction of MC-based IMRT optimization into routine clinical practice, particularly since it has been shown that cost functions converge faster than individual dose uncertainties.[161] For a review of other studies, the reader is referred to the article by Verhaegen and Seuntjens.[25]

III.D.6. Voxel size effects

As is the case with any dose calculation algorithm, calculated dose is affected by the size of the scoring voxel. For MC calculations, typical values in the scoring dimension are voxel sides of 2–5 mm for field sizes greater than 3×3 cm^2 and 1–2 mm for field sizes less than 3×3 cm^2. For calculations where geometric details of the MLC are included in the modeling, scoring voxel sizes no larger than 1–2 mm will be necessary to diminish volume averaging of dose from inter- and intraleaf leakage. As with conventional algorithms, MC-based IMRT calculations should be performed using voxel sizes of 2–3 mm or less in the high gradient regions.[201,202] In addition to affecting the spatial resolution, the statistical uncertainty will be influenced by the voxel size; reducing the voxel size will increase the relative uncertainty for a fixed number of source particles because fewer particles deposit dose in the smaller volume. Increasing the voxel size (and, hence, volume) will reduce the relative uncertainty but may introduce errors due to re-

duced spatial resolution. An example of the influence of volume-averaging effects resulting from the use of larger voxel sizes (~0.5 cm on each side) in MC electron calculations is discussed in Sec. III E 3.5.

III.D.7. Cross sections

The uncertainties in photon interaction cross section data in the energy range from 5 keV to a few MeV is of the order of 1%–2%.[203] Although many MC codes use the incoherent-scattering-factor approximation (which assumes scattering of photons from stationary, free electrons), this approximation is found to be accurate in the mega-voltage energy regime, where the energy of the incident photon is much higher than that of the electron *K*-shell binding energy.[204] An excellent review of the cross sections for bremsstrahlung production and electron-impact ionization has been provided by Seltzer.[205] It is shown that, in general, calculated cross sections for the various electron interaction processes are in agreement with measurements within the combined experimental and theoretical uncertainties.[205] It is felt that cross section and electron data in the megavoltage energy range are sufficiently accurate,[206] assuming that sampling of these cross sections is done accurately within a given code. These same effects will be present in doses computed with the convolution/superposition algorithm.

III.E. Experimental verification

III.E.1. Introduction

In this report experimental verification of the MC algorithm deals with how accurately the algorithm performs under different test conditions within a phantom. As with any algorithm, verification and testing is a necessary step to ensure safety of use in the clinical setting. It is the consensus of this task group that verification of a MC algorithm should be similar to that of any model-based dose calculation algorithm, such as convolution/superposition. The clinical commissioning and acceptance testing of dose calculation algorithms has been reported.[48–50] Additional testing to confirm the accuracy of the MC algorithm in situations of electronic disequilibrium will be helpful. Quantification of benchmark cases (such as the ICCR benchmark[99]) should be performed by either the user, or the vendor. The intent of this section is to provide some examples of additional types of testing that may be included to assure the accuracy of the MC algorithm. The specification of required measurements for acceptance testing and commissioning of the MC algorithm and the criteria for algorithmic agreement with measurements is beyond the scope of this report.

III.E.2. Previous work

Toward the goal of verification of dose calculation algorithms, there have been a variety of studies related to measurements in heterogeneous media. These investigations (see, for example, Refs. 207–212) have usually focused on establishing limitations of conventional dose algorithms in heterogeneous media and on how the use of physics-based algorithms, such as the convolution and MC methods, can be used to produce more accurate results. More recently, Arnfield *et al.*[213] found large differences (up to 10%) between collapsed cone convolution and MC (PEREGRINE) calculations in an 8 cm lung-equivalent slab embedded within solid water and irradiated by a 4×4 cm^2, 18 MV photon beam. Their work included film and ion chamber detectors and showed that MC calculations were in good agreement with measurements.[213]

With respect to patient planning, studies[91,115,214–219] have pointed out major differences between MC calculation and conventional methods, such as the 3D pencil beam and convolution/superposition algorithms. Over the past ten years, there has been a growing interest in the use of the MC method in clinical treatment applications with many institutions around the world actively involved in the development and testing of such systems. Many experiments in homogeneous and heterogeneous phantoms[26,27,71,75,82,84,86,93,96,115,119,194,213,220–225] have been directed toward verification of MC algorithms for clinical planning applications. The interested reader is referred to these and other related publications for a comprehensive review of the various types of experimental testing of MC treatment planning algorithms.

III.E.3. Types of verification experiments

Experimental verification of a MC algorithm should include testing to assess the accuracy of: (a) the beam model (be it measurement-driven or based on treatment head simulation) and (b) the radiation transport algorithm in homogeneous and heterogeneous phantoms. The former is part of routine commissioning of dose calculation algorithms, whereas the latter is likely to have significantly more involvement from developers and vendors.

III.E.3.a. The beam model. The purpose of verification of the beam (treatment head) model is to ensure that parameters, such as the incident beam energy (if used) are correctly "tuned" to produce dose distributions in agreement with measurement. Such verification is the same as that for any dose algorithm and may best be performed with measurements in a homogeneous (water) phantom. These tests should include the acquisition of depth and profile doses in a water phantom for a range of field sizes, as is routinely performed for conventional algorithmic verification, and documented in the AAPM TG-53 report.[48–50] In addition, the use of measured in-air off axis ratios may be useful for benchmarking the beam model. Calculated in-air off-axis ratios have been shown to be very sensitive to the incident electron beam parameters (e.g., mean energy, intensity distribution, etc.), as well as the dimensions and densities of other structures, such as the primary collimator and flattening filter.[102]

The multileaf collimator (MLC). Experiments to benchmark the MLC transport have ranged from arbitrarily shaped AP fields designed to verify overall penumbral and transmission dose[88] to more complicated MLC shapes designed to test modeling of detailed effects, such as transport through the rounded leaf ends,[83,87,226] tongue-and-groove

effect,[27,72,83,87,137] and intra/interleaf transmission.[27,72,83,87,137] The complexity of MLC verification experiments will depend on the usage of the MC algorithm in the clinical setting. For the purposes of 3D conformal radiotherapy (3D-CRT), the shaped field cases provided in Fig. A3–1 of the AAPM TG-53 (Ref. 48) report are good examples of tests designed to test overall MLC transmission and penumbral effects. In the context of 3D-CRT planning, the accurate modeling of details of the MLC and the influence of these effects on leaf transmission and penumbra may be of reduced clinical importance to the target doses (for an appropriate CTV-to-PTV margin) as these issues affect the dose at the field edges and outside the field.[227] For IMRT, it is now well established that the accurate modeling of MLC transmission and penumbra is critical.[61,227] A stringent test of the ability of the MLC model to accurately handle intra/interleaf transmission is presented in Fig. 8(a), for a Varian, Millenium 120-leaf MLC. In this example, the calculated MLC leakage radiation is compared with film measurements in a direction perpendicular to the leaf motion.[83] Figure 8(b) shows a comparison between MC calculations and film measurements for a MLC "picket fence" shape where the field is blocked by even numbered MLC leaves with odd numbered leaves retracted behind the jaws.[83] This test is useful in evaluating how accurately the tongue-and-groove effect is handled.

In designing a test suite for verification of the MLC transport accuracy, the clinical physicist should give special consideration to the detailed specifications of the MLC, particularly if the MC model is to be used for IMRT planning. These details include: density and composition of the MLC leaves, rounded leaf-end dimensions, intra/interleaf gaps, and tongue-and-groove dimensions. Discrepancies between MC calculations and measurements deemed significant by the physicist should be reported to the vendor; modification of the relevant parameters influencing the model accuracy, such as the MLC geometric model should be performed by the vendor to establish acceptable agreement between calculations and measurements.

Other beam modifying devices. The accuracy of MC simulation of beam modifying devices, such as wedges and blocks may be benchmarked using methods similar to that for conventional algorithms.[48] A publication on the experimental verification of the PEREGRINE MC system[27] included the acquisition of large wedged field profiles (40×20 cm^2) at various depths (see Fig. 13 of Hartmann-Siantar *et al.*[27]). The purpose of large field testing is to ensure that radiation transport through the wedge (or other beam modifying device) is in acceptable agreement with measurements across the entire physical range of the device. Unacceptable differences between calculations and measurements should be reported to the vendor—appropriate action should be taken by the vendor to help resolve these differences.

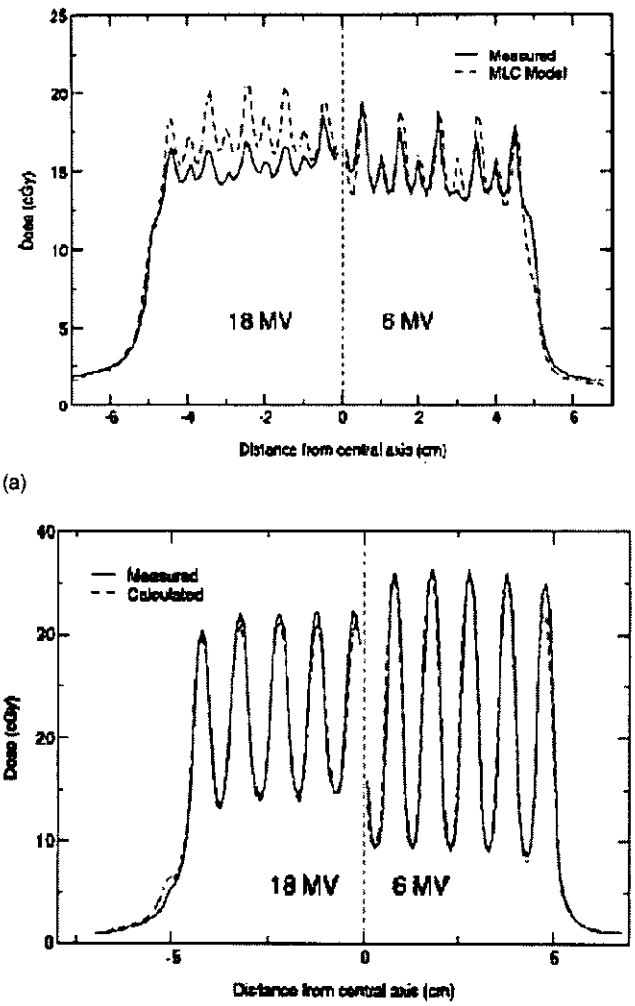


FIG. 8. (a) Measured (solid line) and calculated (dashed line) MLC leakage radiation perpendicular to the direction of MLC leaf motion for a 10 $\times$ 10 cm^2 MLC-blocked field for 6 and 18 MV beams. Dose calculation and measurement for the 6 MV beams occurred at 5 cm depth, 95 cm SSD, and the 18 MV data at 10 cm depth, 90 cm SSD. One should note that the dose due to leakage radiation (under the closed leafs) typically accounts for 2% –3% of the open field dose. The discrepancy noted for the 18 MV comparison is roughly 0.1% of the open field dose and is most likely to a small difference in the MLC density used in the calculations (Ref. 83). Reprinted from Siebers *et al.* (Ref. 83) with permission. (b) Measured (solid line) and calculated (dashed line) doses for 6 and 18 MV 10$\times$10 cm^2 field (a) blocked by even numbered MLC leaves with odd numbered MLC leaves retracted behind the jaws. One should note that the dose due to leakage radiation (under the closed leafs) typically accounts for 2%–3% of the open field dose. Reprinted from Siebers *et al.* (Ref. 83) with permission.

III.E.4. Verification of the Monte Carlo transport algorithm in phantom

Given the improvement in the accuracy of MC simulation over conventional algorithms, particularly under circumstances of electronic disequilibrium, verification of the MC algorithm should include testing under these types of conditions to confirm the expected accuracy. Although reports, such as the AAPM TG-53 (Ref. 48) and others,[49,50] recommend testing in heterogeneous phantoms, issues related to electronic disequilibrium are generally excluded. This task group strongly encourages that verification testing of the MC algorithm include experiments emphasizing electronic disequilibrium effects, in addition to standard tests in heteroge-

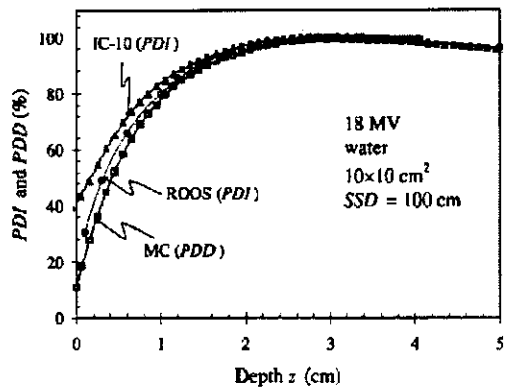


FIG. 9. Measured percent depth ionization and Monte Carlo calculated percent depth dose in the dose buildup region in water with the IC-10 cylindrical ionization chamber and the Roos parallel-plate ionization chamber in water for a 10×10 cm^2 field and SSD of 100 cm, for an 18 MV x-ray beam. For the IC-10 ion chamber, the effective point of measurement was 1.8 mm upstream from the chamber center. Reprinted from Abdel-Rahman *et al.* (Ref. 152) with permission.

neous phantoms as documented in other reports.[48–50] This type of testing is needed to exploit the advantages of MC-based simulation to its full potential.

Verification measurements in slab phantoms with embedded low density inserts and irradiated by high energy photons are useful in assessing the transport accuracy under conditions of electronic disequilibrium.[84,213,220] Effects of penumbral broadening of the dose distribution resulting from lateral electron transport may be evaluated using such phantoms.[84,213,220]

Dose measurements at various locations (depths and off-axis distances) both within and outside a high-density heterogeneity provide a benchmark to evaluate the perturbation of the photon and electron fluence due to the presence of the heterogeneity.[225] Dosimetric effects at interfaces (e.g., due to backscattering at bone/tissue interfaces[53,225]), transport within the heterogeneity, and the doses outside the heterogeneity may all be assessed with careful positioning of measurement devices within the heterogeneous phantom.

Experimental verification should also be performed in more clinically relevant situations. Examples of such geometries include thorax phantoms,[228] and mediastinal[229] and tumor-like phantoms.[229,230]

Verification experiments for algorithmic verification have also included the use of anthropomorphic (Rando) phantoms, where TLDs are most often used for the measurements.[225,231] With anthropomorphic phantoms, calculated patient treatment plans may be verified under a range of clinical circumstances. However, measurements must be carried out carefully with appropriate experimental techniques, theoretical interpretation, and reproducibility.

III.E.5. Dose buildup region

It is challenging to make accurate measurements in the dose buildup region. A recent study by Abdel-Rahman *et al.*[152] showed significant differences (see Fig. 9) in the dose buildup region (<1 cm) between the response of a parallel plate (Roos) and cylindrical ion chambers (IC-10, Scanditronix) in both 6 and 18 MV photon beams. Similar differences were found between a cylindrical ion chamber and a parallel plate chamber (P11) and a stereotactic diode (Scanditronix) by Yokoyama *et al.*[232] who measured dose buildup regions for IMRT fields. Based on the available literature, it is therefore recommended that dose buildups be carefully measured with either a parallel plate chamber, an extrapolation chamber, or with methods such as TLD extrapolation.[233] Diode detectors may also be considered for dose buildup measurements[234] but should be cross referenced with other detectors, such as parallel plate ion chambers. As pointed out by Kawrakow[155] (see Sec. III C 4), the relationship between the measured ionization and dose is sensitive to details of the ion chamber design and needs to be accounted for in dose buildup measurements.

III.E.6. Output ratios

MC calculations of output should be performed in a measurement-like geometry, usually consisting of a central axis point dose estimate at a fixed depth (d_{max}, 10 cm or other) in a water phantom for square field sizes defined by the collimating jaws. If the backscattered radiation into the monitor chamber is correctly modeled, it is possible to calculate output ratios to within 1%–2% agreement with measurements over a range of square field sizes, with sides from 3 to 40 cm.[146] In addition, comprehensive verification of output ratios should include testing for small segmental fields located off-axis, which are commonly used in IMRT planning. Output ratios should be verified against measurements of specially designed IMRT fields, such as junction narrow slit fields, in which the effects of small field dosimetry are magnified. Examples of such fields have been reported.[83,87,145,187]

III.E.7. Electron beams

Careful experimental verification is especially necessary for electrons because the tolerance of parameters, such as the electron energy and the treatment head structure constituents are much tighter than for photons. Accurate measurements are a prerequisite to accurate simulation. Care must be taken in measurement of the central axis depth-dose curve used to define the beam energy. Profiles in large fields need to be included to determine geometry details such as the distance between the scattering foils, the thickness of these foils, and the lateral position of shaped scattering foils, if present. Measurement of dose profiles in the bremsstrahlung tail are helpful to validate the photon component of the beam model.[112] Comparison to measured output ratios and dose distributions over the full clinical range of field sizes and SSDs is necessary for each applicator and beam energy, to rigorously validate the dose calculation. Further measurements, such as dose measured in an anthropomorphic phantom,[221] should be performed to validate the beam model for the specific MC-based application.

Comprehensive verification for electron beam MC simulation should include measurements in heterogeneous phan-

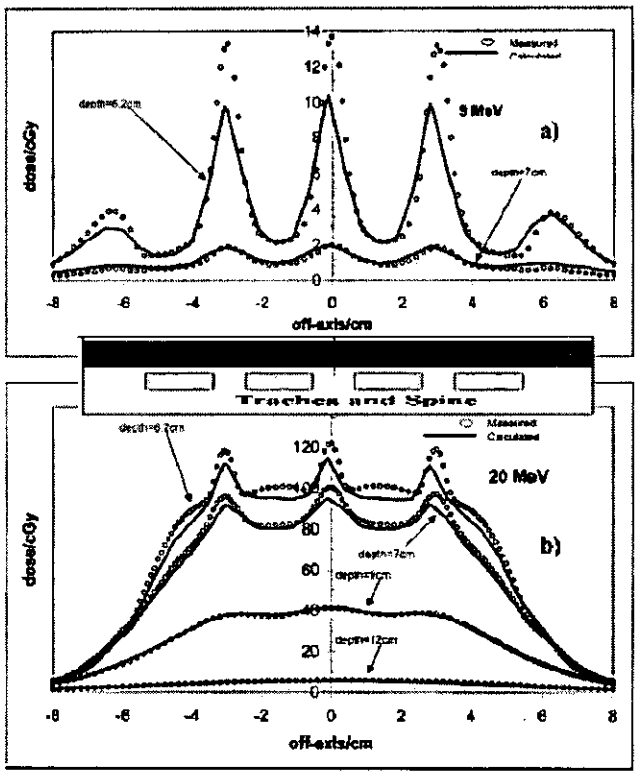


FIG. 10. Trachea and spine phantom: measured and Monte Carlo (Masterplan, Nucletron, based on VMC++) calculated crossplane dose profiles at various depths for: (a) 9 and (b) 20 MeV. SSD=100 cm, 10×10 cm^2 applicator. Monte Carlo simulations were performed with 50 000 histories/cm^2 and a voxel size of 0.49 cm. 100 MU were used for the calculations. The relative uncertainty in the calculations is about 1% −1.5%. The phantom geometry is shown in the inset. Differences between measurements and calculations in this example were attributed to the volume averaging effects of a large voxel size (~0.5 cm) (Ref. 26). Reprinted from Cygler *et al.* (Ref. 26) with permission.

toms, such as those reported in the Electron Collaborative Working Group report[235] and its update.[236] Cygler *et al.*[26] performed experimental benchmarks using a variety of heterogeneous phantoms, including a 1D slab geometry with an aluminum slab insert, a 2D geometry with cortical bone-equivalent inserts ("ribs" geometry), a 3D cylindrical geometry with an "air" insert, and a complex trachea and spine equivalent geometry. Figure 10 shows the example trachea and spine phantom used in their study.[26] Differences between measurements and calculations in this example were shown to be due to the volume averaging effects of a large voxel size (~0.5 cm).[26] The MC calculations are found to underestimate the measurements because the large voxel size averages the density distribution in the vicinity of the heterogeneity, effectively reducing the calculated dose across the low-density heterogeneity.[26]

Careful measurements in such heterogeneous phantoms will provide rigorous benchmark data for verification of the MC algorithm, and as such should be strongly considered for testing purposes.

III.E.8. Measurement uncertainties

Making dose measurements is fraught with problems if the aim is to verify calculations at the 2% or better level, particularly in nonuniform dose regions. It is important to take into account point of measurement effects, variations in stopping power ratios for ion chambers, and beam quality dependence in general for other detectors (e.g., the well known over-response of film in regions of low mean photon energy), polarity effects, ion recombination effects, and chamber perturbation effects. Specific detectors may be required for accurate measurements under specific conditions; for example, properly constructed parallel plate chambers may be favored over ion chambers for measurements in the dose buildup region or at material interfaces. Uncertainties due to improper detector positioning must also be quantified as these measurements are often performed in regions of high dose gradients.

When using modern IMRT techniques there can be strong perturbation effects or corrections needed with ion chamber and other detector measurements. However, as with other algorithms, detector perturbations are generally not accounted for in verification of MC-based treatment planning algorithms. MC methods may be used to model the physical details of the detectors to provide a better understanding of measurements performed under nonequilibrium conditions. These approaches are available to developers of MC techniques but further advancement and investigation of this technology is necessary before specific recommendations can be made. Measurements form the basis for benchmarking the accuracy of dose calculation algorithms in clinical radiotherapy. It is therefore important that the uncertainties in the measuring techniques be accounted for.

III.E.9. Example experimental tests

Example tests for verification of MC dose algorithms are provided in Table IV. These tests, intended as a supplement to those provided in the AAPM TG 53 report,[48] aim to evaluate the various components of the MC algorithm.

III.E.10. Timing issues

The issue of calculation time is of considerable importance in the clinical treatment planning process. To provide a perspective on this subject, users of commercial MC dose algorithms were asked to submit calculation times for photon and electron treatment plans. The treatment plans ranged in complexity from simple AP beams to dynamically delivered IMRT plans. These results are summarized in Table V. As noted in Table V, photon beam calculations with PEREGRINE are approximately a factor of 2–3 times slower than those for conventional planning systems. However, these times will be reduced using faster, currently available processors (>2 GHz versus 800 MHz). Electron beam calculations with VMC++ (Nucletron) are roughly equivalent in processing times to conventional systems. The timing comparisons provide good justification of the fact that significant processing times are no longer a concern for routine clinical MC dose calculation. MC calculations for a plan with one beam take the same time as that for a plan with multiple beams (of approximately the same field size), for the same statistical uncertainty to the PTV. This is because the statistical uncer-

TABLE IV. Partial listing of example specific tests, phantom designs, and detector measurements for Monte Carlo treatment planning systems. These tests are intended as a supplement to those detailed in the AAPM TG-53 (Ref. 48) and other related publications (Refs. 49 and 50) for algorithmic verification. Note that extreme care should be taken when performing many of these measurements as they are, in some instances, highly sensitive to the measurement setup conditions.

Test description	Reason	Phantom design
• Water depth doses and profiles—emphasis on large open field sizes, ($>30\times30$ cm^2) • 2D planar dose perpendicular to the beam cax for large open fields	To evaluate the beam model accuracy—test is sensitive to structures like the flattening filter and other parameters, such as the electron-on-target energy.	• Depth doses and profiles at multiple depths measured in a water phantom using a cylindrical ion chamber.• 2D planar dose at multiple depths in solid water using film.
• 2D planar dose perpendicular to the beam cax of large MLC-shaped fields (see Fig. A3–1 of the AAPM TG 53 report)[a] • Dose profiles under the closed MLC leaves, perpendicular to the direction of motion (see Fig. 7).[b]	To evaluate the accuracy of the MLC model, leaf-tip penumbra and leaf transmission.	• 2D planar dose at multiple depths in solid water using film. • Dose profiles under closed MLC leaves measured with film or small volume detector (diode, TLD, pinpoint chamber, diamond detector).
• Small field (1×1 cm^2–4×4 cm^2) depth doses in low density media; larger field sizes should also be tested. • Penumbral broadening; lateral dose spreading in lung assessed over a range of field sizes (2×2–30×30 cm^2).	To evaluate the transport algorithm accuracy—use of high energies (>10 MV) and low density media emphasizes electronic disequilibrium effects.	• Depth doses in a layered phantom (see Fig. 8 and Fig. 1 of Rice *et al.*)[c] consisting of solid water and low density material (lung equivalent or cork) measured with small volume detector (diode, TLD, pinpoint chamber, diamond detector, at multiple point depths) or with film. Beam is directed perpendicularly to the slabs. • 2D penumbral measurements with film in planes perpendicular to the beam cax at depths above, below, and within the low density slab in the layered phantom. Beam should also be directed parallel to the slabs to evaluate interface effects.[d]
• Depth doses in high density media over a range of field sizes, 3×3–30×30 cm^2.	To evaluate the transport algorithm accuracy in high density media, such as cortical-bone equivalent slabs.	• Depth doses in a layered phantom consisting of solid water and high density material (cortical bone equivalent) measured with small volume detector (diode, TLD, pinpoint chamber, diamond detector) for smaller field sizes, or with film.
• Point doses in the vicinity of tissue interfaces (tissue/lung and tissue/bone), over a range of field sizes, 3×3 –30×30 cm^2.	To evaluate the algorithmic accuracy in the perturbed dose field at tissue interfaces.	See, for example, Fig. 1 of Ref. 230. Dose measured with film or with small volume detector, where possible, at incremental depths, for example, 0.2, 0.5, 1.0, 2.0, and 5.0 cm anterior/posterior to the medial and proximal tissue/lung equivalent and tissue/bone-equivalent interfaces.
• Dose evaluation in clinical treatment planning, for simple, intermediate and complex static treatment plans as well as IMRT plans, in anthropomorphic phantoms.	To assess the accuracy of dose calculation to points located within structures of different densities and receiving different doses based on the treatment plan.	• Dose measured with small volume detectors within inserts of different materials, ranging from air to cortical bone-equivalent. Plans designed should include simple, intermediate, complex static and IMRT beam arrangements. Anthropomorphic phantoms should be CT-imaged for planning purposes.

[a]See Ref. 48.
[b]See Ref. 83.
[c]See Ref. 229.
[d]See Ref. 222.

tainty is determined by the number of particles passing through a volume, and this number can be held constant when performing a MC plan with multiple beams. This is a distinct advantage over conventional algorithms where computational time scales linearly with the number of beams.

IV. CLINICAL IMPLICATIONS OF MONTE CARLO-CALCULATED DOSE DISTRIBUTIONS

IV.A. Introduction

In spite of our confidence in the improved dose calculation accuracy with a suitably commissioned clinical MC algorithm, we are confronted with the following clinical question: What is the effect of more accurate MC dose distributions on patient clinical outcome? To answer this question, we will need to investigate the correlation of MC calculated dose distributions with clinical outcome (in terms of tumor control and normal tissue toxicity).

To date the evidence directly correlating the improved accuracy of MC-calculated dose distributions with clinical outcome is scant. Investigations by De Jaeger *et al.*,[237] (which included convolution but not MC calculations), Chetty *et al.*,[238] and Lindsay *et al.*[239] are among the first

TABLE V. Summary of timing results for clinical treatment plans from currently available commercial Monte Carlo systems. Data for photon beams were provided by author G.E., performed using the PEREGRINE (Nomos division, North American Scientific) system and those for electrons by author J.E.C., conducted with Nucletron. The (1σ) relative statistical uncertainty was approximately 2% in the maximum dose voxel for the Nomos calculations and roughly 1%–1.5% (in the average depth dose along the central axis) for the Nucletron electron beam calculations. Eclipse calculations were reported with an uncertainty of 1%–2% in the mean dose of all voxels receiving more than 50% of the maximum dose within the body of the contour. Readers should be cautioned that the timing results are subject to large uncertainties due to differences in compilers, memory size, cache size, etc.

Monte Carlo code/configuration	Description of treatment plan	Time estimate (min)
PEREGRINE (Nomos, North American Scientific) 16 processors (8-dual), Pentium III, 800 MHz	AP beam, 6 MV photons, 10×10 cm^2 in a water phantom, cubic voxels with 2 mm sides	48
	5 field, 6 MV CRT prostate plan, $\sim7\times11$ cm^2, cubic voxels with ~2.4 mm sides	89
	5 field, 6 MV prostate plan with modulation delivered with DMLC (exposed field $\sim4\times8$ cm^2), cubic voxels with ~2.4 mm sides	71
Masterplan (VMC++, Nucletron), single CPU Pentium IV XEON, 2.2 GHz	AP beam, 6 MeV electrons, 10×10 cm applicator, water phantom, cubic voxels with 4.9 mm sides	4.2
	AP beam, 17 MeV electrons, 10×10 cm applicator, water phantom, cubic voxels with 4.9 mm sides	8.2
	AP beam, 20 MeV electrons, 15×15 cm applicator, water phantom, cubic voxels with 3.9 mm sides	21
	Breast boost treatment plan: 2 fields, 11 MeV electrons, $\sim3\times3$ cm cutouts, cubic voxels with 5.4 mm sides	8.6 (both fields)
eMC (MMC, Eclipse, Varian) single CPU Pentium IV XEON, 2.4 GHz	AP beam, 6, 12, 18 MeV electrons, 10×10 cm applicator, water phantom, cubic voxels with 5.0 mm sides	~3, 4, 4 (6, 12, 18 MeV, respectively)

studies evaluating the influence of improved dose distributions on outcome observed in patients treated with lung cancer. The study by De Jaeger *et al.*,[237] in which lung cancer treatment plans were retrospectively recalculated using a convolution/superposition (CS)-based algorithm (initially calculated with an equivalent-path-length (EPL) algorithm), showed clinically significant differences between calculated and observed incidences of radiation pneumonitis. They demonstrated that the calculated incidence of radiation pneumonitis correlated better with observed incidence when using dose distributions calculated with CS rather than EPL algorithms.[237] Although this study was carried out using a CS algorithm, it provides strong support that the dose-response relationships determined with correction-based algorithms will be different than those computed with model-based methods.[237] With the sometimes large differences observed between the doses calculated with CS and MC algorithms,[240] the MC method is likely to add a higher degree of accuracy to the dose-effect relationships, and will be instrumental in putting these relationships on a more solid footing.

There is clearly a need for more studies addressing the clinical impact of MC-calculated dose distributions. The use of retrospective data may provide a useful means to perform such studies. Retrospective dose assessments of already existing local tumor control and normal tissue complications, using doses recalculated with MC algorithms, may give an early indication of the clinical utility of the MC method, and may also help physicians determine how to use the new MC-calculated doses.[206] Retrospective analyses should eventually show us how to make use of this information in a prospective way.[206]

IV.B. Clinical examples

In reviewing the literature on clinical treatment planning, one should keep in mind that the dose differences found between MC-based and conventional algorithms will be highly dependent on the beam arrangements, field sizes, beam energies, tumor size, and location. This is particularly true in anatomical sites where the target is situated near tissues with widely varying densities, such as the lung and head/neck. For example, due to electron transport issues, differences found in a lung CRT treatment plan using small field sizes and 15 MV photons may be much larger than those found with a standard AP/PA lung plan, using large field sizes and 6 MV photon beams. The reader should therefore be advised that, although there is a general consensus on the importance of the MC method in sites such as the lung, the dosimetry in many of the reported studies is based on specific conditions. The following literature review will focus on treatment planning in the lung and head and neck since differences between MC-based and conventional algorithms are likely to be smaller in other external beam treatment sites. This does not include the potential for improvement in dose estimates in any site from accurate simulation of beam modifiers, in particular, the MLC for delivering IMRT. More studies using MC-based dose calculation techniques in clinical treatment planning are warranted to better quantify the dosimetric and clinical benefits of these algorithms.

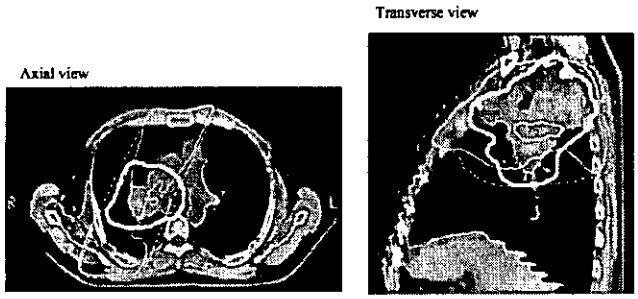


FIG. 11. Opposed, oblique field treatment plan (15 MV photons) showing the 100% isodose coverage for MC (modified DPM, University of Michigan/UMPlan) in the solid line, and an equivalent path length (EPL, University of Michigan/UMPlan) algorithm in the dashed line. The PTV is demarcated in white.

An excellent review of MC-based dose calculation methods in clinical planning for various treatment sites, including the breast and prostate is provided in the article by Reynaert *et al.*[241]

IV.B.1. Photon beam treatment planning

IV.B.1.a. Lung Perhaps the strongest motivation for the need for MC dose calculation comes from treatment planning for lung cancer. This is because electron transport issues, not accounted for accurately with conventional algorithms, are exacerbated in the low density tissues. A consequence of electronic disequilibrium in the lung is the underdosage of the PTV, as shown in Fig. 11. The penumbral widening in the dose distribution as a result of the increased electron scattering in the lung is illustrated for a conformal lung plan in Fig. 12. Depending on the location and size of the tumor, and the beam energy, underdosage of the PTV in lung planning may be significant. In addition to the target coverage, dose to normal tissues, particularly the normal lung, may be equally affected. Numerous lung planning studies have shown sometimes substantial differences (10% −20%) between conventional and MC algorithms.[117,190,192,214,216,242–246]

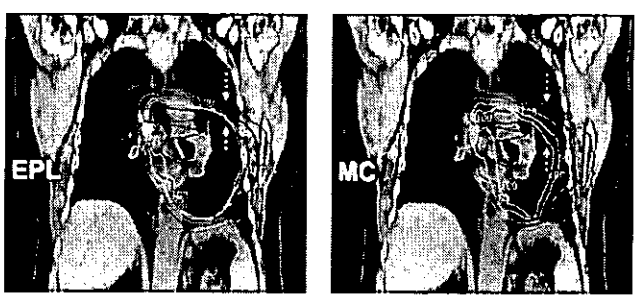


FIG. 12. Isodose distributions for a 3D conformal lung plan (15 MV photons) calculated using an equivalent path length (EPL, University of Michigan/UMPlan) algorithm on the left and a MC calculation (modified DPM, UMPlan) on the right. A distinct penumbral broadening in the MC-based dose distribution (on the right) is observed due to the increased electron scattering in the lung tissue. This effect is not as pronounced in the EPL-calculated dose distribution.

Due to the complicated dosimetric issues associated with treatment planning in the lung, more studies and comparisons using the MC method are encouraged. The utilization of advanced techniques, such as extracranial stereotactic radiotherapy using 10–24 Gy fractions,[247] for the treatment of early stage lung cancer, may increase the importance of accurate dosimetry and the clinical importance of MC calculated doses. One of the important next steps must be to evaluate how the improved MC dose distributions will clinically impact outcome for tumors and normal tissues.

IV.B.1.b. Head and neck There have been numerous studies on MC head and neck CRT and IMRT planning.[190,192,193,214,248–250] These studies have shown in general that dose differences between convolution and MC algorithms are dosimetrically insignificant. However, it has also been shown that dose differences for tumors located in the presence of air cavities can be significant due to inaccurate calculation of the photon/electron energy fluence inside and around air cavities using Batho and ETAR methods[214] as well as a collapsed cone convolution algorithm.[250,251]

IV.B.1.c. Other treatment sites Readers are encouraged to review articles published on MC-based treatment planning in other anatomical sites, which include: brain,[89,91,219,252–254] breast,[216,255] and prostate.[193,214,216,256,257]

IV.B.2. Electron beam treatment planning

There have been numerous studies on the use of MC calculations in electron beam treatment planning.[26,111,115,121,258–262] In some instances, significant differences between pencil beam dose distributions and MC calculations have been demonstrated, particularly in regions in or near air cavities and lung or bone tissues.[115,263] Large differences have also been observed for small irregular fields, beams with oblique incidence, and for extended SSD treatments.[115,263] The use of the MC method represents a significant improvement in accuracy for electron beam dose calculation compared with conventional algorithms. Electron MC calculations require far fewer primary histories to achieve a given uncertainty on the dose being calculated because electrons deposit their energy in a more continuous manner, in a smaller volume and mostly starting from the same location (the patient surface). This timing advantage has allowed the development of commercial electron beam MC algorithms which are time efficient (taking on the order of minutes per plan) even on a single processor.[26]

From phantom and routine patient planning calculations based on experience with a commercial electron beam MC algorithm (VMC++, currently Nucletron), Cygler *et al.*[26] have reported that MC-generated monitor unit (MU) values may differ from the homogeneous (water phantom) values by as much as 10%, depending on surface irregularities and inhomogeneities present in the irradiated volume. This situation is frequently encountered for head and neck sites, especially in the region of the nasal cavities. In such cases, surface irregularities and missing tissue lateral to the air cavity necessitate larger MUs to deliver the same prescribed dose versus flat surface anatomies. As is the case with photon beams,

it is important that the statistical uncertainties in the single voxel MU calculations be reported.

IV.C. Association of Monte Carlo calculated dose distributions with clinical outcome

As with other changes to the therapy treatment process, with the implementation of a new dose calculation algorithm such as the MC method, users should correlate doses and prescriptions with respect to previous clinical experience.

Chetty *et al.*[238] studied dose-effect relationships (for tumor and normal tissues) by recalculating dose distributions retrospectively using the MC method for patients treated on a nonsmall cell lung cancer (NSCLC) dose escalation protocol[264]—original plans were generated using an EPL algorithm. Follow-up data was evaluated from CT scans taken six months to two years postradiation therapy. Follow-up CT scans were fused with initial treatment planning scans and the original and replanned dose distributions were mapped onto the anatomy to establish associations between dose and regions of local recurrence and normal lung damage (radiation-induced pneumonitis).[238] Preliminary results of this study showed that the originally planned PTVs are sometimes significantly underdosed with MC calculations compared with the EPL algorithm.[238] For normal lung tissue, the correlation of dose with normal tissue complications was also found to differ, but also showed that beam model differences (not related to the dose calculation algorithm in the patient) are important and must be considered for an unbiased comparison of dose calculated by different algorithms.[136,238,241,265]

Lindsay *et al.*[239] performed retrospective MC-based recalculations of a large group of lung cancer treatment plans and showed significant differences in dose indices (V20, maximum lung dose and mean GTV dose) between plans without heterogeneity correction and MC calculations. Moreover, correlations between V20 and observed radiation pneumonitis in this study were found to be different between plans without heterogeneity correction and MC-based treatment plans.[239]

Despite the preliminary (and somewhat anecdotal) nature of the evidence thus far, observations suggest that more accurate dose calculations will reveal inadequate target coverage or hot spots in certain areas of organs at risk that could lead to differences in outcome. Complete MC recalculations of delivered dose distributions that include the effects of other factors, such as organ motion and patient setup errors, will be required to refine these correlation studies. In addition, it should be noted that the efficacy of radiation therapy is also dependent upon individual patient response.

The clinical evidence thus far, albeit preliminary and retrospective, provides support that dose delivery based on MC treatment plans, particularly for lung cancer, have the potential to result in clinically significant changes. Kong *et al.*[266] have shown that dose significantly impacts local control and overall survival for NSCLC; local control was found to increase at a rate of 1.3% per gray above the conventional dose fractionation scheme (63–69 Gy in 2.0 Gy fractions). This suggests that even small differences in dose distributions, as a result of inaccurate dose calculation, are likely to affect local control and survival for patients with NSCLC. Further studies of MC dose distributions in the lung and in other sites, such as the head/neck, are necessary and are encouraged in order to unequivocally evaluate the clinical utility of MC-calculated doses.

V. SUMMARY

We wish to reiterate that, from a clinical implementation standpoint, the MC method should be treated as would any conventional dose algorithm. Proper implementation will require the clinical physicist to understand, on some level, the fundamentals of the algorithm as well as the possible pitfalls associated with its clinical implementation, much as one would for any dose algorithm. In addition to providing an educational review of the MC dose calculation method, we have identified issues that will need to be considered by developers, vendors, and end users of MC-based techniques, ultimately to ensure that patient dose calculations are effected safely.

In conclusion, we present a summary of some of the issues which are important in the implementation and clinical use of MC dose calculation algorithms. The recommendations summarized here are not meant to be prescriptive; rather they are intended as a preliminary guide for medical physicists to ensure thoughtful and safe implementation of clinical MC algorithms. We anticipate that many of the areas of concern will be studied in further detail in future, more specific task group reports.

V.A. Treatment head simulation

(a) Vendors of MC-based dose calculation systems should be responsible for providing the necessary guidance and assistance with the beam modeling and benchmarking process. Such guidance includes the tuning of parameters such as the electron energy, or the adjustment of model parameters (in measurement-driven models) to ensure that the beam model meets the required specifications.

(b) In reporting statistical uncertainties in the calculated patient dose, vendors should properly account for latent variance in the beam model if the model is based on treatment head simulation. Further, if the latent variance is a significant contributor to the total variance the number of phase-space particles in the beam simulation should be increased.

V.B. Patient simulation

V.B.1. Statistical uncertainties

We recommend that quantities, such as $F_{D>0.5D_{max}}$, or F_{PTV}, or F_{PRV} for doses to the specific volumes, PTV or PRV respectively, be adopted as a standard method of reporting fractional statistical uncertainties in dose averaged over the relevant volume. The sole use of dose uncertainties to indi-

vidual voxels, such as the maximum dose voxel, D_{max}, should be avoided. Additionally, reporting of single voxel doses or doses over patient subvolumes should be accompanied by their respective statistical uncertainties. In situations where doses in individual voxels are important, such as D_{max} to a serial organ like the spinal cord, it may be necessary to simulate a large enough number of histories so that $s_{D_{max}}$ is very small. This will ensure that the absolute uncertainty in D_{max} will also be small. The required statistical precision required for individual voxel dose estimates should be decided upon with guidance from the clinical team.

V.B.2. Variance reduction techniques, efficiency enhancing methods, and other parameters

Users should understand the influence on the dose accuracy of variance reduction implementations and approximate methods used to improve the calculation efficiency, as well as any other accessible parameters of importance to the MC dose algorithm. Appropriate documentation on the methods used and their influence should be made available to the user. More studies on the influence of efficiency enhancing methods in clinical treatment planning are warranted, as it is likely that the tradeoffs between speed and accuracy will not be the same in different anatomical sites.[267] Where possible, vendors should provide users with the flexibility to adjust parameter inputs for these efficiency enhancement techniques but implement default values which are conservative, i.e., accurate in all situations.

V.B.3. Dose prescriptions

Vendors are strongly discouraged from using single voxel (point) doses for dose prescription and monitor unit calculations in their MC-based treatment planning systems. Rather, doses should be prescribed to volumes larger than a single voxel, such as the volume contained by an isodose surface. Current users of MC-based planning systems are encouraged to find ways of circumventing point-based dose prescriptions if their systems are not flexible enough to allow otherwise.

V.B.4. CT-to-material conversions

The use of conversion techniques based purely on mass density (i.e., assuming the only patient material is water, but with varying density), as employed in conventional algorithms, is discouraged with MC simulation because these methods ignore dependencies of particle interactions on the materials, which can lead to notable discrepancies in high atomic number materials. The conversions should include the use of both mass density and the atomic compositions of the materials. Appropriate documentation on the CT number-to-material conversion method used by the software should be accessible to the user.

V.B.5. Dose-to-water and dose-to-medium

MC dose results should: (a) explicitly indicate the material to which the dose is computed, (b) allow conversion between D_m and D_w using the methods discussed in Sec. III D 4, or other methods as developed in future investigations. It is strongly encouraged that appropriate documentation on the dose-to-water conversion method used by the software be provided to the user.

V.C. Experimental verification

V.C.1. Examples of specific tests

In addition to the standard testing necessary for conventional dose algorithms, it is strongly recommended that additional tests (e.g., as suggested in Sec. III E 3) be included to evaluate the accuracy of MC calculations under situations where the MC method is known to perform better than conventional algorithms.

V.C.2. Verification calculations

MC verification calculations should be performed under the same conditions as the experiments. Phantoms used should be CT-imaged for planning purposes. Statistical uncertainties for verification plans should be reported and included in the comparison of calculations with measurements. More studies of the influence of detector perturbations using MC calculations are encouraged.

V.C.3. Measurement uncertainties

It is important for the user to understand the limitations of the various measurement devices as they are used in different situations. It is recommended that realistic measurement uncertainties be assessed and included when evaluating MC verification calculations against measurements.

Although the implementation of MC treatment planning will require clinical physicists to once again understand a new technology, one should view this in a positive light. Properly implemented MC algorithms will provide dose calculation with sufficient accuracy (in at least a single instance of the patient geometry) such that dose calculation is no longer a source of meaningful uncertainty in the radiotherapy planning process. In addition, although the details are sometimes complex, the underlying idea of simulating the actual transport of individual particles is simpler to understand than many other algorithms because it is based on actual physical processes familiar to the medical physicist. Finally, this may well be the last new dose calculation algorithm that medical physicists will need to learn since MC techniques provide the highest level of accuracy for dose calculations in radiotherapy treatment planning.

ACKNOWLEDGMENTS

I.J.C. acknowledges the support of his colleagues, in particular, Dick Fraass, Daniel McShan, and Randy Ten Haken, at the University of Michigan, Department of Radiation Oncology. Thanks to Neelam Tyagi and Mihaela Rosu for their help with compiling references. We are grateful to Lesley Buckley, Dan La Russa, and Randy Taylor of Carleton University for commenting on early versions of the manuscript. I.J.C. is thankful to Alex Bielajew for helpful discussions on various topics in this report. I.J.C. acknowledges the support

of NIH Grant Nos. R01CA106770 (UNMC) and P01CA59827 (previously at the University of Michigan). D.W.O.R. is supported by the Canada Research Chairs program and the Natural Sciences and Engineering Research Council of Canada. B.A.F. acknowledges the support of NIH Grant No. R01CA104777. J.V.S. acknowledges the support of NIH Grant No. R01CA98524. J. S. is a research scientist of the National Cancer Institute Canada supported with funds from the Canadian Cancer Society. Finally, we are indebted to Alan Nahum for his careful review of this report which helped us improve the pedagogical aspects and other topics significantly.

[a]Electronic mail: ichetty@unmc.edu

[1]N. Papanikolaou, J. Battista, A. Boyer, C. Kappas, E. Klein, T. Mackie, M. Sharpe, and J. Van Dyk, "AAPM Report No. 85: Tissue inhomogeneity corrections for megavoltage photon beams," in *AAPM Report No. 85* (Medical Physics, Madison, WI, 2004), pp. 1–135.

[2]A. Dutreix, "When and how can we improve precision in radiotherapy?," Radiother. Oncol. **2**, 275–292 (1984).

[3]C. G. Orton, P. M. Mondalek, J. T. Spicka, D. S. Herron, and L. I. Andres, "Lung corrections in photon beam treatment planning: Are we ready?," Int. J. Radiat. Oncol. Biol. Phys. **10**, 2191–2199 (1984).

[4]J. G. Stewart and A. W. Jackson, "The steepness of the dose response curve both for tumor cure and normal tissue injury," Laryngoscope **85**, 1107–1111 (1975).

[5]M. Goitein and J. Busse, "Immobilization error: Some theoretical considerations," Radiology **117**, 407–412 (1975).

[6]http://eom.springer.de/B/b017750.htm

[7]N. Metropolis, "The beginning of the MC Method," Los Alamos Sci. **15**, 125–130 (1987); see http://library.lanl.gov/la-pubs/00326866.pdf.

[8]D. W. O. Rogers and A. F. Bielajew, in *The Dosimetry of Ionizing Radiation*, edited by B. Bjarngard, K. Kase, and F. Attix (Academic, New York, 1990), Vol. III, pp. 427–539.

[9]M. J. Berger, in *Methods in Computational Physics*, edited by S. Fernbach, B. Alder, and M. Rothenberg (Academic, New York, 1963), Vol. 1.

[10]M. Berger and S. Seltzer, "ETRAN Monte Carlo code system for electron and photon transport through extended media," Radiation Shielding Information Center (RSIC) Report CCC-107, Oak Ridge National Laboratory, Oak Ridge, TN, 1973.

[11]J. F. Briesmeister, "MCNP—A general Monte Carlo *N*-particle transport code, version 4A," Report LA-12625-M, Los Alamos National Laboratory, Los Alamos, NM, 1993.

[12]W. R. Nelson, H. Hirayama, and D. W. O. Rogers, "The EGS4 code system," Report SLAC-265, Stanford Linear Accelerator, Stanford, CA, 1985.

[13]P. L. Petti, M. S. Goodman, T. A. Gabriel, and R. Mohan, "Investigation of buildup dose from electron contamination of clinical photon beams," Med. Phys. **10**, 18–24 (1983).

[14]R. Mohan, C. Chui, and L. Lidofsky, "Energy and angular distributions of photons from medical linear accelerators," Med. Phys. **12**, 592–597 (1985).

[15]M. Udale, "A Monte Carlo investigation of surface doses for broad electron beams," Phys. Med. Biol. **33**, 939–954 (1988).

[16]R. L. Ford and W. R. Nelson, "The EGS code system—Version 3," Report SLAC-210, Stanford Linear Accelerator, Stanford, CA, 1978.

[17]I. Kawrakow and D. W. O. Rogers, "The EGSnrc code system: Monte Carlo simulation of electron and photon transport," Technical Report PIRS-701, National Research Council of Canada, Ottawa, Ontario, 2000.

[18]P. R. Almond, P. J. Biggs, B. M. Coursey, W. F. Hanson, M. S. Huq, R. Nath, and D. W. O. Rogers, "AAPM's TG-51 protocol for clinical reference dosimetry of high-energy photon and electron beams," Med. Phys. **26**, 1847–1870 (1999).

[19]B. A. Faddegon, P. O'Brien, and D. L. Mason, "The flatness of Siemens linear accelerator x-ray fields," Med. Phys. **26**, 220–228 (1999).

[20]M. B. Tacke, H. Szymanowski, U. Oelfke, C. Schulze, S. Nuss, E. Wehrwein, and S. Leidenberger, "Assessment of a new multileaf collimator concept using GEANT4 Monte Carlo simulations," Med. Phys. **33**, 1125–1132 (2006).

[21]A. Ito, in *Monte Carlo Transport of Electrons and Photons*, edited by W. R. Nelson, T. M. Jenkins, A. Rindi, A. E. Nahum, and D. W. O. Rogers (Plenum, New York, 1988), pp. 573–598.

[22]A. E. Nahum, in *Monte Carlo Transport of Electrons and Photons*, edited by W. R. Nelson, T. M. Jenkins, A. Rindi, A. E. Nahum, and D. W. O. Rogers (Plenum, New York, 1988), pp. 3–20.

[23]D. W. O. Rogers, B. A. Faddegon, G. X. Ding, C. M. Ma, J. We, and T. R. Mackie, "BEAM: A Monte Carlo code to simulate radiotherapy treatment units," Med. Phys. **22**, 503–524 (1995).

[24]C.-M. Ma and S. B. Jiang, "Monte Carlo modelling of electron beams from medical accelerators," Phys. Med. Biol. **44**, R157–R189 (1999).

[25]F. Verhaegen and J. Seuntjens, "Monte Carlo modelling of external radiotherapy photon beams," Phys. Med. Biol. **48**, R107–R164 (2003).

[26]J. E. Cygler, G. M. Daskalov, G. H. Chan, and G. X. Ding, "Evaluation of the first commercial Monte Carlo dose calculation engine for electron beam treatment planning," Med. Phys. **31**, 142–153 (2004).

[27]C. L. Hartmann Siantar, "Description and dosimetric verification of the PEREGRINE Monte Carlo dose calculation system for photon beams incident on a water phantom," Med. Phys. **28**, 1322–1337 (2001).

[28]P. Andreo, "Monte Carlo techniques in medical radiation physics," Phys. Med. Biol. **36**, 861–920 (1991).

[29]D. E. Raeside, "Monte Carlo principles and applications," Phys. Med. Biol. **21**, 181–197 (1976).

[30]D. W. O. Rogers, "Fifty years of Monte Carlo simulations for medical physics," Phys. Med. Biol. **51**, R287–R301 (2006).

[31]J. E. Turner, H. A. Wright, and R. N. Hamm, "A Monte Carlo primer for health physicists," Health Phys. **48**, 717–733 (1985).

[32]F. B. Brown, "MCNP—A general Monte Carlo-particle transport code, version 5," Report LA-UR-03 1987, Los Alamos National Laboratory, Los Alamos, NM, 2003.

[33]S. Agostinelli, "GEANT4—A simulation toolkit," Nucl. Instrum. Methods Phys. Res. A **506**, 250–303 (2003).

[34]J. Baro, J. Sempau, J. M. Fernandez-Varea, and F. Salvat, "PENELOPE—An algorithm for Monte-Carlo simulation of the penetration and energy-loss of electrons and positrons in matter," Nucl. Instrum. Methods Phys. Res. A **100**, 31–46 (1995).

[35]M. Fippel, "Fast Monte Carlo dose calculation for photon beams based on the VMC electron algorithm," Med. Phys. **26**, 1466–1475 (1999).

[36]I. Kawrakow, "VMC++ , electron and photon Monte Carlo calculations optimized for radiation treatment planning," in *Advanced Monte Carlo for Radiation Physics, Particle Transport Simulation and Applications: Proceedings of the Monte Carlo 2000 Meeting Lisbon*, edited by A. Kling, F. Barao, M. Nakagawa, L. Tavora, and P. Vaz (Springer, Berlin, 2001), pp. 229–236.

[37]I. Kawrakow, M. Fippel, and K. Friedrich, "3D electron dose calculation using a Voxel based Monte Carlo algorithm (VMC)," Med. Phys. **23**, 445–457 (1996).

[38]H. Neuenschwander and E. J. Born, "A macro Monte-Carlo method for electron-beam dose calculations," Phys. Med. Biol. **37**, 107–125 (1992).

[39]J. Sempau, S. J. Wilderman, and A. F. Bielajew, "DPM, a fast, accurate Monte Carlo code optimized for photon and electron radiotherapy treatment planning dose calculations," Phys. Med. Biol. **45**, 2263–2291 (2000).

[40]I. Kawrakow, "Accurate condensed history Monte Carlo simulation of electron transport. I. EGSnrc, the new EGS4 version," Med. Phys. **27**, 485–498 (2000).

[41]A. F. Bielajew and D. W. O. Rogers, in *Monte Carlo Transport of Electrons and Photons*, edited by W. R. Nelson, T. M. Jenkins, A. Rindi, A. E. Nahum, and D. W. O. Rogers (Plenum, New York, 1988), pp. 115–137.

[42]I. Kawrakow and A. F. Bielajew, "On the condensed history technique for electron transport," Nucl. Instrum. Methods Phys. Res. B **142**, 253–280 (1998).

[43]E. W. Larsen, "A theoretical derivation of the condensed history algorithm," Ann. Nucl. Energy **19**, 701–714 (1992).

[44]I. Kawrakow, "Accurate condensed history Monte Carlo simulation of electron transport. II. Application to ion chamber response simulations," Med. Phys. **27**, 499–513 (2000).

[45]I. Kawrakow and M. Fippel, "Investigation of variance reduction techniques for Monte Carlo photon dose calculation using XVMC," Phys. Med. Biol. **45**, 2163–2183 (2000).

[46]I. Kawrakow, D. W. O. Rogers, and B. R. B. Walters, "Large efficiency improvements in BEAMnrc using directional bremsstrahlung splitting," Med. Phys. **31**, 2883–2898 (2004).

[47]I. Kawrakow and B. R. B. Walters, "Efficient photon beam dose calculations using DOSXYZnrc with BEAMnrc," Med. Phys. **33**, 3046–3056 (2006).
[48]B. Fraass, K. Doppke, M. Hunt, G. Kutcher, G. Starkschall, R. Stern, and J. Van Dyke, "American Association of Physicists in Medicine Radiation Therapy Committee Task Group 53: Quality assurance for clinical radiotherapy treatment planning," Med. Phys. **25**, 1773–1829 (1998).
[49]IAEA-Technical Report Series No. 430: Commissioning and quality qssurance of computerized planning systems for radiation treatment of cancer," in International Atomic Energy Agency, Vienna, 2004.
[50]J. Van Dyk, R. B. Barnett, J. E. Cygler, and P. C. Shragge, "Commissioning and quality assurance of treatment planning computers," Int. J. Radiat. Oncol. Biol. Phys. **26**, 261–273 (1993).
[51]D. W. O. Rogers and A. F. Bielajew, in *Monte Carlo Transport of Electrons and Photons*, edited by W. R. Nelson, T. M. Jenkins, A. Rindi, A. E. Nahum, and D. W. O. Rogers (Plenum, New York, 1988), pp. 407–419.
[52]I. Kawrakow, "On the efficiency of photon beam treatment head simulations," Med. Phys. **32**, 2320–2326 (2005).
[53]C.-M. Ma *et al.*, "A Monte Carlo dose calculation tool for radiotherapy treatment planning," Phys. Med. Biol. **47**, 1671–1689 (2002).
[54]J. A. Halbleib, in *Monte Carlo Transport of Electrons and Photons*, edited by W. R. Nelson, T. M. Jenkins, A. Rindi, A. E. Nahum, and D. W. O. Rogers (Plenum, New York, 1988), pp. 249–262.
[55]J. A. Halbleib and T. A. Melhorn, "ITS: The integrated TIGER series of coupled electron/photon Monte Carlo transport codes," Sandia Report SAND84–0573, Sandia National Laboratory, Albuquerque, NM, 1984.
[56]S. M. Seltzer, in *Monte Carlo Transport of Electrons and Photons*, edited by W. R. Nelson, T. M. Jenkins, A. Rindi, A. E. Nahum, and D. W. O. Rogers (Plenum, New York, 1988), pp. 153–182.
[57]D. W. O. Rogers, B. Walters, and I. Kawrakow, "BEAMnrc Users Manual," NRC Report PIRS 509(a)revH, 2004.
[58]B. R. B. Walters and D. W. O. Rogers, "DOSXYZnrc Users Manual," NRC Report PIRS 794 (rev B), 2004.
[59]D. W. O. Rogers, "The role of Monte-Carlo simulation of electron-transport in radiation-dosimetry," Appl. Radiat. Isot. **42**, 965–974 (1991).
[60]T. R. Mackie, S. S. Kubsad, D. W. O. Rogers, and A. F. Bielajew, "The OMEGA project: Electron dose planning using Monte Carlo simulation," Med. Phys. **17**, 730 (abstract) (1990).
[61]J. Siebers and R. Mohan, "Monte Carlo and IMRT," in *Intensity Modulated Radiation Therapy, The State of the Art, Proceedings of the 2003 AAPM Summer School*, edited by T. R. Mackie and J. R. Palta (Advanced Medical, Madison, WI, 2003), pp. 531–560.
[62]A. Leal, F. Sanchez-Doblado, R. Arrans, M. Perucha, M. Rincon, E. Carrasco, and C. Bernal, "Monte Carlo simulation of complex radiotherapy treatments," Comput. Sci. Eng. **6**, 60–68 (2004).
[63]N. Tyagi, A. Bose, and I. J. Chetty, "Implementation of the DPM Monte Carlo Code on a parallel architecture for treatment planning applications," Med. Phys. **31**, 2721–2725 (2004).
[64]J. M. Fernandez-Varea, R. Mayol, J. Baro, and F. Salvat, "On the theory and simulation of multiple elastic-scattering of electrons," Nucl. Instrum. Methods Phys. Res. B **73**, 447–473 (1993).
[65]A. E. Schach von Wittenau, L. J. Cox, P. M. Bergstrom, Jr., W. P. Chandler, C. L. Hartmann Siantar, and R. Mohan, "Correlated histogram representation of Monte Carlo derived medical accelerator photon-output phase space," Med. Phys. **26**, 1196–1211 (1999).
[66]A. E. Schach von Wittenau, P. M. Bergstrom, Jr., and L. J. Cox, "Patient-dependent beam-modifier physics in Monte Carlo photon dose calculations," Med. Phys. **27**, 935–947 (2000).
[67]I. Kawrakow, "Electron transport: Multiple and plural scattering," Nucl. Instrum. Methods Phys. Res. B **108**, 23–34 (1996).
[68]I. Kawrakow and A. F. Bielajew, "Recent improvements and accuracy tests of the VOXEL Monte Carlo algorithm," Med. Phys. **24**, 1049 (abstract) (1997).
[69]P. J. Keall and P. W. Hoban, "Superposition dose calculation incorporating Monte Carlo generated electron track kernels," Med. Phys. **23**, 479–485 (1996).
[70]L. Wang, C. S. Chui, and M. Lovelock, "A patient-specific Monte Carlo dose-calculation method for photon beams," Med. Phys. **25**, 867–878 (1998).
[71]R. Doucet, M. Olivares, F. DeBlois, E. B. Podgorsak, I. Kawrakow, and J. Seuntjens, "Comparison of measured and Monte Carlo calculated dose distributions in inhomogeneous phantoms in clinical electron beams," Phys. Med. Biol. **48**, 2339–2354 (2003).
[72]M. Fippel, "Efficient particle transport simulation through beam modulating devices for Monte Carlo treatment planning," Med. Phys. **31**, 1235–1242 (2004).
[73]M. Fippel, F. Haryanto, O. Dohm, F. Nusslin, and S. Kriesen, "A virtual photon energy fluence model for Monte Carlo dose calculation," Med. Phys. **30**, 301–311 (2003).
[74]M. Fippel, I. Kawrakow, and K. Friedrich, "Electron beam dose calculations with the VMC algorithm and the verification data of the NCI working group," Phys. Med. Biol. **42**, 501–520 (1997).
[75]M. Fippel, W. Laub, B. Huber, and F. Nusslin, "Experimental investigation of a fast Monte Carlo photon beam dose calculation algorithm," Phys. Med. Biol. **44**, 3039–3054 (1999).
[76]M. Fippel and F. Nusslin, "Evaluation of a clinical Monte Carlo dose calculation code based on the ICCR benchmark test," Med. Phys. **28**, 1198 (abstract) (2001).
[77]I. Kawrakow and A. F. Bielajew, "On the representation of electron multiple elastic-scattering distributions for Monte Carlo calculations," Nucl. Instrum. Methods Phys. Res. B **134**, 325–336 (1998).
[78]H. Neuenschwander, T. R. Mackie, and P. J. Reckwerdt, "MMC—A high-performance Monte Carlo code for electron beam treatment planning," Phys. Med. Biol. **40**, 543–574 (1995).
[79]C. Cris, E. Born, R. Mini, H. Neuenschwander, and W. Volken, "A scaling method for multiple source models," in *Proceedings of the 13th ICCR*, edited by T. Bortfeld and W. Schlegel (Springer-Verlag, Heidelberg, 2000), pp. 411–413.
[80]P. Pemler, J. Besserer, U. Schneider, and H. Neuenschwander, "Evaluation of a commercial electron treatment planning system based on Monte Carlo techniques (eMC)," Z. Med. Phys. **16**, 313–329 (2006).
[81]C.-M. Ma, J. S. Li, T. Pawlicki, S. B. Jiang, and J. Deng, "MCDOSE—A Monte Carlo dose calculation tool for radiation therapy treatment planning," in *Proceedings of the 13th ICCR*, edited by T. Bortfeld and W. Schlegel (Springer-Verlag, Heidelberg, 2000), pp. 411–413.
[82]J. V. Siebers, P. J. Keall, J. Kim, and R. Mohan, "Performance benchmarks of the MCV Monte Carlo System," in *Proceedings of the 13th ICCR*, edited by T. Bortfeld and W. Schlegel (Springer-Verlag, Heidelberg, 2000), pp. 129–131.
[83]J. V. Siebers, P. J. Keall, J. O. Kim, and R. Mohan, "A method for photon beam Monte Carlo multileaf collimator particle transport," Phys. Med. Biol. **47**, 3225–3249 (2002).
[84]I. J. Chetty, P. M. Charland, N. Tyagi, D. L. McShan, B. A. Fraass, and A. F. Bielajew, "Photon beam relative dose validation of the DPM Monte Carlo code in lung-equivalent media," Med. Phys. **30**, 563–573 (2003).
[85]I. J. Chetty, N. Tyagi, M. Rosu, P. M. Charland, D. L. McShan, R. K. Ten Haken, B. A. Fraass, and A. F. Bielajew, "Clinical implementation, validation and use of the DPM Monte Carlo code for radiotherapy treatment planning," in *Nuclear Mathematical and Computational Sciences: A Century in Review, A Century Anew, Gatlinburg, TN* (American Nuclear Society, LaGrange Park, IL, 2003), Vol. 119, pp. 1–17.
[86]J. J. DeMarco, T. D. Solberg, and J. B. Smathers, "A CT-based Monte Carlo simulation tool for dosimetry planning and analysis," Med. Phys. **25**, 1–11 (1998).
[87]R. F. Aaronson, J. J. DeMarco, I. J. Chetty, and T. D. Solberg, "A Monte Carlo based phase space model for quality assurance of intensity modulated radiotherapy incorporating leaf specific characteristics," Med. Phys. **29**, 2952–2958 (2002).
[88]I. Chetty, J. J. DeMarco, and T. D. Solberg, "A virtual source model for Monte Carlo modeling of arbitrary intensity distributions," Med. Phys. **27**, 166–172 (2000).
[89]I. J. Chetty, J. J. DeMarco, T. D. Solberg, A. R. Arellano, R. Fogg, and A. V. Mesa, "A phase space model for simulating arbitrary intensity distributions for shaped radiosurgery beams using the Monte Carlo method," Radiosurgery **3**, 41–52 (2000).
[90]T. D. Solberg *et al.*, "A review of radiation dosimetry applications using the MCNP Monte Carlo code," Radiochim. Acta **89**, 337–355 (2001).
[91]T. D. Solberg, J. J. DeMarco, F. E. Holly, J. B. Smathers, and A. A. F. DeSalles, "Monte Carlo treatment planning for stereotactic radiosurgery," Radiother. Oncol. **49**, 73–84 (1998).
[92]M. K. Fix, P. Manser, E. J. Born, R. Mini, and P. Ruegsegger, "Monte Carlo simulation of a dynamic MLC based on a multiple source model," Phys. Med. Biol. **46**, 3241–3257 (2001).
[93]M. K. Fix, M. Stampanoni, P. Manser, E. J. Born, R. Mini, and P. Ruegsegger, "A multiple source model for 6 MV photon beam dose calculations using Monte Carlo," Phys. Med. Biol. **46**, 1407–1427 (2001).

[94]E. Poon and F. Verhaegen, "Accuracy of the photon and electron physics in GEANT4 for radiotherapy applications," Med. Phys. **32**, 1696–1711 (2005).
[95]J. Sempau, A. Sanchez-Reyes, and F. Salvat, "H. O. ben Tahar, S. B. Jiang, and J. M. Fernandez-Varea, "Monte Carlo simulation of electron beams from an accelerator head using PENELOPE," Phys. Med. Biol. **46**, 1163–1186 (2001).
[96]O. Chibani and X. A. Li, "Monte Carlo dose calculations in homogeneous media and at interfaces: A comparison between GEPTS, EGSnrc, MCNP, and measurements," Med. Phys. **29**, 835–847 (2002).
[97]O. Chibani and C. M. Ma, "Electron depth dose distributions in water, iron and lead: The GEPTS system," Nucl. Instrum. Methods Phys. Res. B **101**, 357–378 (1995).
[98]W. van der Zee, A. Hogenbirk, and S. C. van der Marck, "ORANGE: A Monte Carlo dose engine for radiotherapy," Phys. Med. Biol. **50**, 625–641 (2005).
[99]D. W. O. Rogers and R. Mohan, "Questions for comparisons of clinical Monte Carlo codes," in *Proceedings of the 13th ICCR*, edited by T. Bortfeld and W. Schlegel (Springer-Verlag, Heidelberg, 2000), pp. 120–122.
[100]B. A. Faddegon and I. Blevis, "Electron spectra derived from depth dose distributions," Med. Phys. **27**, 514–526 (2000).
[101]M. R. Bieda, J. A. Antolak, and K. R. Hogstrom, "The effect of scattering foil parameters on electron-beam Monte Carlo calculations," Med. Phys. **28**, 2527–2534 (2001).
[102]D. Sheikh-Bagheri and D. W. O. Rogers, "Sensitivity of megavoltage photon beam Monte Carlo simulations to electron beam and other parameters," Med. Phys. **29**, 379–390 (2002).
[103]A. Tzedakis, J. E. Damilakis, M. Mazonakis, J. Stratakis, H. Varveris, and N. Gourtsoyiannis, "Influence of initial electron beam parameters on Monte Carlo calculated absorbed dose distributions for radiotherapy photon beams," Med. Phys. **31**, 907–913 (2004).
[104]I. J. Chetty, P. M. Charland, N. Tyagi, D. L. McShan, B. Fraass, and A. F. Bielajew, "Experimental validation of the DPM Monte Carlo code for photon beam dose calculations in inhomogeneous media," Med. Phys. **29**, 1351 (abstract) (2002).
[105]B. Libby, J. Siebers, and R. Mohan, "Validation of Monte Carlo generated phase-space descriptions of medical linear accelerators," Med. Phys. **26**, 1476–1483 (1999).
[106]C. Bramoulle, F. Husson, and J. P. Manens, "Monte Carlo (PENELOPE code) study of the x-ray beams from SL Linacs (Elekta)," Phys. Med. **16**, 107–115 (2000).
[107]E. L. Chaney, T. J. Cullip, and T. A. Gabriel, "A Monte Carlo study of accelerator head scatter," Med. Phys. **21**, 1383–1390 (1994).
[108]G. X. Ding, "Energy spectra, angular spread, fluence profiles and dose distributions of 6 and 18 MV photon beams: Results of Monte Carlo simulations for a Varian 2100EX accelerator," Phys. Med. Biol. **47**, 1025–1046 (2002).
[109]R. A. C. Siochi, "Requirements for manufacturer supplied data for Monte Carlo simulation," in *Proceedings of the 15th International Conference on the Applications of Accelerators in Research and Industry* (The American Institute of Physics, Melville, 1999), pp. 1060–1065.
[110]G. G. Zhang, D. W. O. Rogers, J. E. Cygler, and T. R. Mackie, "Monte Carlo investigation of electron beam output factors versus size of square cutout," Med. Phys. **26**, 743–750 (1999).
[111]J. A. Antolak, M. R. Bieda, and K. R. Hogstrom, "Using Monte Carlo methods to commission electron beams: A feasibility study," Med. Phys. **29**, 771–786 (2002).
[112]B. Faddegon, E. Schreiber, and X. Ding, "Monte Carlo simulation of large electron fields," Phys. Med. Biol. **50**, 741–753 (2005).
[113]E. C. Schreiber and B. A. Faddegon, "Sensitivity of large-field electron beams to variations in a Monte Carlo accelerator model," Phys. Med. Biol. **50**, 769–778 (2005).
[114]C.-M. Ma, B. A. Faddegon, D. W. O. Rogers, and T. R. Mackie, "Accurate characterization of Monte Carlo calculated electron beams for radiotherapy," Med. Phys. **24**, 401–416 (1997).
[115]C.-M. Ma, E. Mok, A. Kapur, T. Pawlicki, D. Findley, S. Brain, K. Forster, and A. L. Boyer, "Clinical implementation of a Monte Carlo treatment planning system," Med. Phys. **26**, 2133–2143 (1999).
[116]C.-M. Ma and D. W. O. Rogers, "BEAM Characterization: A Multiple-Source Model," NRC Report PIRS-0509(C), 1995.
[117]C.-M. Ma, "Characterization of computer simulated radiotherapy beams for Monte-Carlo treatment planning," Radiat. Phys. Chem. **53**, 329–344 (1998).
[118]J. Deng, S. B. Jiang, A. Kapur, J. Li, T. Pawlicki, and C. M. Ma, "Photon beam characterization and modelling for Monte Carlo treatment planning," Phys. Med. Biol. **45**, 411–427 (2000).
[119]B. Faddegon, J. Balogh, R. Mackenzie, and D. Scora, "Clinical considerations of Monte Carlo for electron radiotherapy treatment planning," Radiat. Phys. Chem. **53**, 217–227 (1998).
[120]M. K. Fix, H. Keller, P. Ruegsegger, and E. J. Born, "Simple beam models for Monte Carlo photon beam dose calculations in radiotherapy," Med. Phys. **27**, 2739–2747 (2000).
[121]S. B. Jiang, A. Kapur, and C. M. Ma, "Electron beam modeling and commissioning for Monte Carlo treatment planning," Med. Phys. **27**, 180–191 (2000).
[122]S. B. Jiang, J. Deng, J. Li, P. Pawlicki, A. Boyer, and C.-M. Ma, "Modeling and commissioning of clinical photon beams for Monte Carlo treatment planning," in *Proceedings of the 13th ICCR*, edited by T. Bortfeld and W. Schlegel (Springer-Verlag, Heidelberg, 2000), pp. 434–436.
[123]J. Deng, S. B. Jiang, P. Pawlicki, J. Li, and C.-M. Ma, "Electron beam commissioning for Monte Carlo dose calculation," in *Proceedings of the 13th ICCR*, edited by T. Bortfeld and W. Schlegel (Springer-Verlag, Heidelberg, 2000), pp. 431–433.
[124]J. S. Li *et al.*, "Source modeling and beam commissioning for Siemens photon beams," Med. Phys. **29**, 1230 (abstract) (2002).
[125]J. Yang, J. S. Li, L. Qin, W. Xiong, and C. M. Ma, "Modelling of electron contamination in clinical photon beams for Monte Carlo dose calculation," Phys. Med. Biol. **49**, 2657–2673 (2004).
[126]J. Deng, T. Guerrero, C. M. Ma, and R. Nath, "Modelling 6 MV photon beams of a stereotactic radiosurgery system for Monte Carlo treatment planning," Phys. Med. Biol. **49**, 1689–1704 (2004).
[127]J. Deng, S. B. Jiang, T. Pawlicki, J. Li, and C. M. Ma, "Derivation of electron and photon energy spectra from electron beam central axis depth dose curves," Phys. Med. Biol. **46**, 1429–1449 (2001).
[128]J. J. Janssen, E. W. Korevaar, L. J. van Battum, P. R. Storchi, and H. Huizenga, "A model to determine the initial phase space of a clinical electron beam from measured beam data," Phys. Med. Biol. **46**, 269–286 (2001).
[129]S. Siljamaki, L. Tillikainen, H. Helminen, and J. Pyyry, "Determining parameters for a multiple-source model of a linear accelerator using optimization techniques," Med. Phys. **32**, 2113 (abstract) (2005).
[130]W. Ulmer, J. Pyyry, and W. Kaissl, "A 3D photon superposition/convolution algorithm and its foundation on results of Monte Carlo calculations," Phys. Med. Biol. **50**, 1767–1790 (2005).
[131]K. Aljarrah, G. C. Sharp, T. Neicu, and S. B. Jiang, "Determination of the initial beam parameters in Monte Carlo linac simulation," Med. Phys. **33**, 850–858 (2006).
[132]A. Ahnesjo and P. Andreo, "Determination of effective bremsstrahlung spectra and electron contamination for photon dose calculations," Phys. Med. Biol. **34**, 1451–1464 (1989).
[133]A. Ahnesjo and A. Trepp, "Acquisition of the effective lateral energy fluence distribution for photon beam dose calculations by convolution models," Phys. Med. Biol. **36**, 973–985 (1991).
[134]A. Ahnesjo, L. Weber, A. Murman, M. Saxner, I. Thorslund, and E. Traneus, "Beam modeling and verification of a photon beam multisource model," Med. Phys. **32**, 1722–1737 (2005).
[135]A. Catala, P. Francois, J. Bonnet, and C. Scouarnec, "Reconstruction of 12 MV bremsstrahlung spectra from measured transmission data by direct resolution of the numeric system $AF=T$," Med. Phys. **22**, 3–10 (1995).
[136]P. M. Charland, I. J. Chetty, L. D. Paniak, B. P. Bednarz, and B. A. Fraass, "Enhanced spectral discrimination through the exploitation of interface effects in photon dose data," Med. Phys. **31**, 264–276 (2004).
[137]E. Heath and J. Seuntjens, "Development and validation of a BEAMnrc component module for accurate Monte Carlo modelling of the Varian dynamic millennium multileaf collimator," Phys. Med. Biol. **48**, 4045–4063 (2003).
[138]H. H. Liu, F. Verhaegen, and L. Dong, "A method of simulating dynamic multileaf collimators using Monte Carlo techniques for intensity-modulated radiation therapy," Phys. Med. Biol. **46**, 2283–2298 (2001).
[139]N. Tyagi, J. M. Moran, D. W. Litzenberg, A. F. Bielajew, B. A. Fraass, and I. J. Chetty, "Experimental verification of a Monte Carlo-based MLC simulation model for IMRT dose calculation," Med. Phys. **34**, 651–663 (2007).
[140]T. C. Zhu *et al.*, "Output ratios in air: Report of the AAPM Task Group No. 74," Med. Phys. (submitted).
[141]T. C. Zhu and B. E. Bjarngard, "Head scatter off-axis for megavoltage x

rays," Med. Phys. **30**, 533–543 (2003).
[142]H. H. Liu, T. R. Mackie, and E. C. McCullough, "A dual source photon beam model used in convolution/superposition dose calculations for clinical megavoltage x-ray beams," Med. Phys. **24**, 1960–1974 (1997).
[143]M. B. Sharpe, D. A. Jaffray, J. J. Battista, and P. Munro, "Extrafocal radiation: A unified approach to the prediction of beam penumbra and output factors for megavoltage x-ray beams," Med. Phys. **22**, 2065–2074 (1995).
[144]H. H. Liu, T. R. Mackie, and E. C. McCullough, "Calculating output factors for photon beam radiotherapy using a convolution/superposition method based on a dual source photon beam model," Med. Phys. **24**, 1975–1985 (1997).
[145]M. R. Arnfield, J. V. Siebers, J. O. Kim, Q. Wu, P. J. Keall, and R. Mohan, "A method for determining multileaf collimator transmission and scatter for dynamic intensity modulated radiotherapy," Med. Phys. **27**, 2231–2241 (2000).
[146]H. H. Liu, T. R. Mackie, and E. C. McCullough, "Modeling photon output caused by backscattered radiation into the monitor chamber from collimator jaws using a Monte Carlo technique," Med. Phys. **27**, 737–744 (2000).
[147]S. B. Jiang, A. L. Boyer, and C. M. Ma, "Modeling the extrafocal radiation and monitor chamber backscatter for photon beam dose calculation," Med. Phys. **28**, 55–66 (2001).
[148]G. X. Ding, "Using Monte Carlo simulations to commission photon beam output factors—A feasibility study," Phys. Med. Biol. **48**, 3865–3874 (2003).
[149]B. Parker, A. S. Shiu, and H. H. Liu, "Small-field dosimetry with multiple detectors and Monte Carlo calculations," Med. Phys. **29**, 1372 (abstract) (2002).
[150]G. X. Ding, "Dose discrepancies between Monte Carlo calculations and measurements in the buildup region for a high-energy photon beam," Med. Phys. **29**, 2459–2463 (2002).
[151]G. X. Ding, C. Duzenli, and N. I. Kalach, "Are neutrons responsible for the dose discrepancies between Monte Carlo calculations and measurements in the build-up region for a high-energy photon beam?" Phys. Med. Biol. **47**, 3251–3261 (2002).
[152]W. Abdel-Rahman, J. P. Seuntjens, F. Verhaegen, F. Deblois, and E. B. Podgorsak, "Validation of Monte Carlo calculated surface doses for megavoltage photon beams," Med. Phys. **32**, 286–298 (2005).
[153]I. Kawrakow, "Efficient photon beam treatment head simulations," Radiother. Oncol. **81**, 82 (abstract) (2006).
[154]D. Sheikh-Bagheri, D. W. O. Rogers, C. K. Ross, and J. P. Seuntjens, "Comparison of measured and Monte Carlo calculated dose distributions from the NRC linac," Med. Phys. **27**, 2256–2266 (2000).
[155]I. Kawrakow, "On the effective point of measurement in megavoltage photon beams," Med. Phys. **33**, 1829–1839 (2006).
[156]O. Chibani and C. M. Ma, "On the discrepancies between Monte Carlo dose calculations and measurements for the 18 MV varian photon beam," Med. Phys. **34**, 1206–1216 (2007).
[157]W. Feller, *An Introduction to Probability Theory and Its Applications*, 3rd ed. (Wiley, New York, 1967), Vol. I.
[158]B. R. B. Walters, I. Kawrakow, and D. W. O. Rogers, "History by history statistical estimators in the BEAM code system," Med. Phys. **29**, 2745–2752 (2002).
[159]J. Sempau and A. F. Bielajew, "Towards the elimination of Monte Carlo statistical fluctuation from dose volume histograms for radiotherapy treatment planning," Phys. Med. Biol. **45**, 131–157 (2000).
[160]P. J. Keall, J. V. Siebers, R. Jeraj, and R. Mohan, "The effect of dose calculation uncertainty on the evaluation of radiotherapy plans," Med. Phys. **27**, 478–484 (2000).
[161]I. Kawrakow, "The effect of Monte Carlo statistical uncertainties on the evaluation of dose distributions in radiation treatment planning," Phys. Med. Biol. **49**, 1549–1556 (2004).
[162]I. J. Chetty, M. Rosu, M. L. Kessler, B. A. Fraass, R. K. Ten Haken, F. M. Kong, and D. L. McShan, "Reporting and analyzing statistical uncertainties in Monte Carlo-based treatment planning," Int. J. Radiat. Oncol. Biol. Phys. **65**, 1249–1259 (2006).
[163]J. V. Siebers, P. J. Keall, and I. Kawrakow, in *The Modern Technology of Radiation Oncology*, edited by J. Van Dyke (Medical Physics, Madison, WI, 2005), Vol. 2, pp. 91–130.
[164]F. M. Buffa and A. E. Nahum, "Monte Carlo dose calculations and radiobiological modelling: Analysis of the effect of the statistical noise of the dose distribution on the probability of tumour control," Phys. Med. Biol. **45**, 3009–3023 (2000).
[165]S. B. Jiang, T. Pawlicki, and C. M. Ma, "Removing the effect of statistical uncertainty on dose-volume histograms from Monte Carlo dose calculations," Phys. Med. Biol. **45**, 2151–2161 (2000).
[166]J. O. Deasy, "Denoising of electron beam Monte Carlo dose distributions using digital filtering techniques," Phys. Med. Biol. **45**, 1765–1779 (2000).
[167]I. Kawrakow, "On the de-noising of Monte Carlo calculated dose distributions," Phys. Med. Biol. **47**, 3087–3103 (2002).
[168]J. O. Deasy, M. V. Wickerhauser, and M. Picard, "Accelerating Monte Carlo simulations of radiation therapy dose distributions using wavelet threshold de-noising," Med. Phys. **29**, 2366–2373 (2002).
[169]S. J. Pollack and A. F. Bielajew, "Novel algorithms for smoothing global Monte Carlo noise," in Proceedings of the Current Topics in Monte Carlo Treatment Planning: Advanced Workshop, Montreal, CN, edited by F. Verhaegen and J. Seuntjens, 2004 (unpublished).
[170]B. Miao, R. Jeraj, S. Bao, and T. R. Mackie, "Adaptive anisotropic diffusion filtering of Monte Carlo dose distributions," Phys. Med. Biol. **48**, 2767–2781 (2003).
[171]M. Fippel and F. Nusslin, "Smoothing Monte Carlo calculated dose distributions by iterative reduction of noise," Phys. Med. Biol. **48**, 1289–1304 (2003).
[172]I. El Naqa *et al.*, "A comparison of Monte Carlo dose calculation denoising techniques," Phys. Med. Biol. **50**, 909–922 (2005).
[173]C. M. Ma, R. A. Price, Jr., J. S. Li, L. Chen, L. Wang, E. Fourkal, L. Qin, and J. Yang, "Monitor unit calculation for Monte Carlo treatment planning," Phys. Med. Biol. **49**, 1671–1687 (2004).
[174]F. C. du Plessis, C. A. Willemse, M. G. Lotter, and L. Goedhals, "The indirect use of CT numbers to establish material properties needed for Monte Carlo calculation of dose distributions in patients," Med. Phys. **25**, 1195–1201 (1998).
[175]C.-M. Ma and D. W. O. Rogers, "BEAMDP Users Manual," NRC Report PIRS-0509(D), 1995.
[176]J. V. Siebers, P. J. Keall, A. E. Nahum, and R. Mohan, "Converting absorbed dose to medium to absorbed dose to water for Monte Carlo based photon beam dose calculations," Phys. Med. Biol. **45**, 983–995 (2000).
[177]ICRU-Report No. 46: Photon, electron, proton and neutron interaction data for body tissues," in International Commission on Radiation Units and Measurements, 1992.
[178]F. Verhaegen and S. Devic, "Sensitivity study for CT image use in Monte Carlo treatment planning," Phys. Med. Biol. **50**, 937–946 (2005).
[179]M. Bazalova, L. Beaulieu, S. Palefsky, and F. Verhaegen, "Correction of CT artifacts and its influence on Monte Carlo dose calculations," Med. Phys. **34**, 2119–2132 (2007).
[180]C. Reft *et al.*, "Dosimetric considerations for patients with HIP prostheses undergoing pelvic irradiation. Report of the AAPM Radiation Therapy Committee Task Group 63," Med. Phys. **30**, 1162–1182 (2003).
[181]N. Dogan, J. V. Siebers, and P. J. Keall, "Clinical comparison of head and neck and prostate IMRT plans using absorbed dose to medium and absorbed dose to water," Phys. Med. Biol. **51**, 4967–4980 (2006).
[182]H. H. Liu, "D_m rather than D_w should be used in Monte Carlo treatment planning. For the proposition," Med. Phys. **29**, 922–923 (2002).
[183]M. Goitein, "The cell's-eye view: assessing dose in four dimensions," Int. J. Radiat. Oncol. Biol. Phys. **62**, 951–953 (2005).
[184]M. Fippel and F. Nusslin, "Comments on 'Converting absorbed dose to medium to absorbed dose to water for Monte Carlo based photon beam dose calculations'," Phys. Med. Biol. **45**, L17–L19 (2000).
[185]J. Siebers, B. Libby, and R. Mohan, "Trust, but verify: Comparison of MCNP and BEAM Monte Carlo codes for generation of phase space distributions for a Varian 2100C," Med. Phys. **25**, A143 (abstract) (1998).
[186]R. Mohan, M. Arnfield, S. Tong, Q. Wu, and J. Siebers, "The impact of fluctuations in intensity patterns on the number of monitor units and the quality and accuracy of intensity modulated radiotherapy," Med. Phys. **27**, 1226–1237 (2000).
[187]J. O. Kim, J. V. Siebers, P. J. Keall, M. R. Arnfield, and R. Mohan, "A Monte Carlo study of radiation transport through multileaf collimators," Med. Phys. **28**, 2497–2506 (2001).
[188]P. J. Keall, J. V. Siebers, M. Arnfield, J. O. Kim, and R. Mohan, "Monte Carlo dose calculations for dynamic IMRT treatments," Phys. Med. Biol. **46**, 929–941 (2001).
[189]J. V. Siebers, M. Lauterbach, P. J. Keall, and R. Mohan, "Incorporating multi-leaf collimator leaf sequencing into iterative IMRT optimization,"

Med. Phys. **29**, 952–959 (2002).
[190]L. Wang, E. Yorke, and C. S. Chui, "Monte Carlo evaluation of 6 MV intensity modulated radiotherapy plans for head and neck and lung treatments," Med. Phys. **29**, 2705–2717 (2002).
[191]C.-M. Ma *et al.*, "Monte Carlo verification of IMRT dose distributions from a commercial treatment planning optimization system," Phys. Med. Biol. **45**, 2483–2495 (2000).
[192]T. Pawlicki and C. M. Ma, "Monte Carlo simulation for MLC-based intensity-modulated radiotherapy," Med. Dosim. **26**, 157–168 (2001).
[193]P. Francescon, S. Cora, and P. Chiovati, "Dose verification of an IMRT treatment planning system with the BEAM EGS4-based Monte Carlo code," Med. Phys. **30**, 144–157 (2003).
[194]M. Rincon *et al.*, "Monte Carlo conformal treatment planning as an independent assessment," in *Advanced Monte Carlo for Radiation Physics: Proceedings of the Monte Carlo 2000 Meeting, Lisbon*, edited by A. Kling *et al.* (Springer-Verlag, Berlin, 2001), pp. 565–570.
[195]R. Jeraj, P. J. Keall, and J. V. Siebers, "The effect of dose calculation accuracy on inverse treatment planning," Phys. Med. Biol. **47**, 391–407 (2002).
[196]N. Reynaert *et al.*, "The importance of accurate linear accelerator head modelling for IMRT Monte Carlo calculations," Phys. Med. Biol. **50**, 831–846 (2005).
[197]W. Laub, M. Alber, M. Birkner, and F. Nusslin, "Monte Carlo dose computation for IMRT optimization," Phys. Med. Biol. **45**, 1741–1754 (2000).
[198]N. Dogan, J. V. Siebers, P. J. Keall, F. Lerma, Y. Wu, M. Fatyga, J. F. Williamson, and R. K. Schmidt-Ullrich, "Improving IMRT dose accuracy via deliverable Monte Carlo optimization for the treatment of head and neck cancers," Med. Phys. **33**, 4033–4055 (2006).
[199]J. V. Siebers, M. Lauterbach, S. Tong, Q. Wu, and R. Mohan, "Reducing dose calculation time for accurate iterative IMRT planning," Med. Phys. **29**, 231–237 (2002).
[200]A. M. Bergman, K. Bush, M. P. Milette, I. A. Popescu, K. Otto, and C. Duzenli, "Direct aperture optimization for IMRT using Monte Carlo generated beamlets," Med. Phys. **33**, 3666–3679 (2006).
[201]B. De Smedt, B. Vanderstraeten, N. Reynaert, W. De Neve, and H. Thierens, "Investigation of geometrical and scoring grid resolution for Monte Carlo dose calculations for IMRT," Phys. Med. Biol. **50**, 4005–4019 (2005).
[202]J. F. Dempsey, H. E. Romeijn, J. G. Li, D. A. Low, and J. R. Palta, "A Fourier analysis of the dose grid resolution required for accurate IMRT fluence map optimization," Med. Phys. **32**, 380–388 (2005).
[203]J. H. Hubbell, "Review of photon interaction cross section data in the medical and biological context," Phys. Med. Biol. **44**, R1–R22 (1999).
[204]D. V. Rao, S. M. Seltzer, and P. M. Bergstrom, Jr., "Compton scattering cross-sections for individual subshells for a few elements of biological interest in the energy region 5 keV–10 MeV," Radiat. Phys. Chem. **70**, 479–489 (2004).
[205]S. M. Seltzer, in *Monte Carlo Transport of Electrons and Photons*, edited by W. R. Nelson, T. M. Jenkins, A. Rindi, A. E. Nahum, and D. W. O. Rogers (Plenum, New York, 1988), pp. 81–114.
[206]B. A. Fraass, J. Smathers, and J. Deye, "Summary and recommendations of a National Cancer Institute workshop on issues limiting the clinical use of Monte Carlo dose calculation algorithms for megavoltage external beam radiation therapy," Med. Phys. **30**, 3206–3216 (2003).
[207]E. R. Epp, A. L. Boyer, and K. P. Doppke, "Underdosing of lesions resulting from lack of electronic equilibrium in upper respiratory air cavities irradiated by 10 MV x-ray beams," Int. J. Radiat. Oncol. Biol. Phys. **2**, 613–619 (1977).
[208]M. A. Hunt, G. E. Desobry, B. Fowble, and L. R. Coia, "Effect of low-density lateral interfaces on soft-tissue doses," Int. J. Radiat. Oncol. Biol. Phys. **37**, 475–482 (1997).
[209]E. E. Klein, L. M. Chin, R. K. Rice, and B. J. Mijnheer, "The influence of air cavities on interface doses for photon beams," Int. J. Radiat. Oncol. Biol. Phys. **27**, 419–427 (1993).
[210]T. R. Mackie, J. W. Scrimger, and J. J. Battista, "A convolution method of calculating dose for 15-MV x rays," Med. Phys. **12**, 188–196 (1985).
[211]R. Mohan, C. Chui, and L. Lidofsky, "Differential pencil beam dose computation model for photons," Med. Phys. **13**, 64–73 (1986).
[212]C. X. Yu, J. W. Wong, and J. A. Purdy, "Photon dose perturbations due to small inhomogeneities," Med. Phys. **14**, 78–83 (1987).
[213]M. R. Arnfield, C. H. Siantar, J. Siebers, P. Garmon, L. Cox, and R. Mohan, "The impact of electron transport on the accuracy of computed dose," Med. Phys. **27**, 1266–1274 (2000).
[214]F. C. du Plessis, C. A. Willemse, M. G. Lotter, and L. Goedhals, "Comparison of the Batho, ETAR and Monte Carlo dose calculation methods in CT based patient models," Med. Phys. **28**, 582–589 (2001).
[215]A. O. Jones and I. J. Das, "Comparison of inhomogeneity correction algorithms in small photon fields," Med. Phys. **32**, 766–776 (2005).
[216]T. Knoos, E. Wieslander, L. Cozzi, C. Brink, A. Fogliata, D. Albers, H. Nystrom, and S. Lassen, "Comparison of dose calculation algorithms for treatment planning in external photon beam therapy for clinical situations," Phys. Med. Biol. **51**, 5785–5807 (2006).
[217]M. Miften, M. Wiesmeyer, A. Kapur, and C. M. Ma, "Comparison of RTP dose distributions in heterogeneous phantoms with the BEAM Monte Carlo simulation system," J. Appl. Clin. Med. Phys. **2**, 21–31 (2001).
[218]R. Mohan, "Why Monte Carlo?," in *Proceedings of the 12th ICCR*, edited by D. Leavitt (Medical Physics, Salt Lake City, UT, 1997), pp. 16–18.
[219]S. N. Rustgi, A. K. Rustgi, S. B. Jiang, and K. M. Ayyangar, "Dose perturbation caused by high-density inhomogeneities in small beams in stereotactic radiosurgery," Phys. Med. Biol. **43**, 3509–3518 (1998).
[220]P. Carrasco *et al.*, "Comparison of dose calculation algorithms in phantoms with lung equivalent heterogeneities under conditions of lateral electronic disequilibrium," Med. Phys. **31**, 2899–2911 (2004).
[221]J. Coleman, C. Joy, J. E. Park, P. Villarreal-Barajas, P. L. Petti, and B. Faddegon, "A comparison of Monte Carlo and Fermi–Eyges–Hogstrom estimates of heart and lung dose from breast electron boost treatment," Int. J. Radiat. Oncol. Biol. Phys. **61**, 621–628 (2005).
[222]T. Krieger and O. A. Sauer, "Monte Carlo-versus pencil-beam-/collapsed-cone dose calculation in a heterogeneous multi-layer phantom," Phys. Med. Biol. **50**, 859–868 (2005).
[223]W. U. Laub, A. Bakai, and F. Nusslin, "Intensity modulated irradiation of a thorax phantom: Comparisons between measurements, Monte Carlo calculations and pencil beam calculations," Phys. Med. Biol. **46**, 1695–1706 (2001).
[224]E. Spezi, D. G. Lewis, and C. W. Smith, "Monte Carlo simulation and dosimetric verification of radiotherapy beam modifiers," Phys. Med. Biol. **46**, 3007–3029 (2001).
[225]L. Wang, M. Lovelock, and C. S. Chui, "Experimental verification of a CT-based Monte Carlo dose-calculation method in heterogeneous phantoms," Med. Phys. **26**, 2626–2634 (1999).
[226]K. De Vlamynck, H. Palmans, F. Verhaegen, C. De Wagter, W. De Neve, and H. Thierens, "Dose measurements compared with Monte Carlo simulations of narrow 6 MV multileaf collimator shaped photon beams," Med. Phys. **26**, 1874–1882 (1999).
[227]G. A. Ezzell *et al.*, "Guidance document on delivery, treatment planning, and clinical implementation of IMRT: Report of the IMRT Subcommittee of the AAPM Radiation Therapy Committee," Med. Phys. **30**, 2089–2115 (2003).
[228]C. G. Orton, P. M. Mondalek, J. T. Spicka, D. S. Herron, and L. I. Andres, "Benchmark measurements for lung dose corrections for x-ray beams," Int. J. Radiat. Oncol. Biol. Phys. **10**, 2191–2199 (1984).
[229]R. K. Rice, B. J. Mijnheer, and L. M. Chin, "Benchmark measurements for lung dose corrections for x-ray-beams," Int. J. Radiat. Oncol. Biol. Phys. **15**, 399–409 (1988).
[230]P. M. Charland, I. J. Chetty, S. Yokoyama, and B. A. Fraass, "Dosimetric comparison of extended dose range film with ionization measurements in water and lung equivalent heterogeneous media exposed to megavoltage photons," J. Appl. Clin. Med. Phys. **4**, 25–39 (2003).
[231]P. Dunscombe, P. McGhee, and E. Lederer, "Anthropomorphic phantom measurements for the validation of a treatment planning system," Phys. Med. Biol. **41**, 399–411 (1996).
[232]S. Yokoyama, P. L. Roberson, D. L. Litzenberg, J. M. Moran, and B. A. Fraass, "Surface buildup dose dependence on photon field delivery technique for IMRT," J. Appl. Clin. Med. Phys. **5**, 71–81 (2004).
[233]T. Kron, A. Elliot, T. Wong, G. Showell, B. Clubb, and P. Metcalfe, "X-ray surface dose measurements using TLD extrapolation," Med. Phys. **20**, 703–711 (1993).
[234]B. E. Bjarngard, P. Vadash, and T. Zhu, "Doses near the surface in high-energy x-ray beams," Med. Phys. **22**, 465–468 (1995).
[235]A. S. Shiu *et al.*, "Verification data for electron beam dose algorithms," Med. Phys. **19**, 623–636 (1992).
[236]R. A. Boyd, K. R. Hogstrom, J. A. Antolak, and A. S. Shiu, "A measured data set for evaluating electron-beam dose algorithms," Med. Phys. **28**, 950–958 (2001).
[237]K. De Jaeger, M. S. Hoogeman, M. Engelsman, Y. Seppenwoolde, E. M.

F. Damen, B. J. Mijnheer, L. J. Boersma, and J. V. Lebesque, "Incorporating an improved dose-calculation algorithm in conformal radiotherapy of lung cancer: Re-evaluation of dose in normal lung tissue," Radiother. Oncol. **69**, 1–10 (2003).

[238]I. J. Chetty, M. Rosu, F.-M. Kong, C. Lopez, D. S. Tatro, D. L. McShan, B. A. Fraass, and R. K. Ten Haken, "On the correlation of dose-volume-response using Monte Carlo dose calculation in conformal radiation therapy of lung cancer," in *Proceedings of the 14th ICCR*, edited by B.Y. Yi, S. D. Ahn, E. K. Choi, and S. W. Ha (Jeong, Seoul, Korea, 2004), pp. 457–460.

[239]P. E. Lindsay, I. El Naqa, A. Hope, M. Vicic, J. Cui, J. Bradley, and J. O. Deasy, "Retrospective Monte Carlo dose calculations with limited beam weight information," Med. Phys. **34**, 334–346 (2007).

[240]P. J. Keall, J. Siebers, and R. Mohan, "The impact of Monte Carlo dose calculations on treatment outcomes," in *Proceedings of the 13th ICCR*, edited by T. Bortfeld and W. Schlegel (Springer-Verlag, Heidelberg, 2000), pp. 425–427.

[241]N. Reynaert *et al.*, "Monte Carlo treatment planning for photon and electron beams," Radiat. Phys. Chem. **76**, 643–686 (2007).

[242]A. Fogliata, E. Vanetti, D. Albers, C. Brink, A. Clivio, T. Knoos, G. Nicolini, and L. Cozzi, "On the dosimetric behaviour of photon dose calculation algorithms in the presence of simple geometric heterogeneities: Comparison with Monte Carlo calculations," Phys. Med. Biol. **52**, 1363–1385 (2007).

[243]T. Knoos, A. Ahnesjo, P. Nilsson, and L. Weber, "Limitations of a pencil beam approach to photon dose calculations in lung tissue," Phys. Med. Biol. **40**, 1411–1420 (1995).

[244]P. N. McDermott, T. He, and A. DeYoung, "Dose calculation accuracy of lung planning with a commercial IMRT treatment planning system," J. Appl. Clin. Med. Phys. **4**, 341–351 (2003).

[245]M. F. Tsiakalos, K. Theodorou, C. Kappas, S. Zefkili, and J. C. Rosenwold, "Analysis of the penumbra enlargement in lung versus the quality index of photon beams: A methodology to check the dose calculation algorithm," Med. Phys. **31**, 943–949 (2004).

[246]E. D. Yorke, L. Wang, K. E. Rosenzweig, D. Mah, J. B. Paoli, and C. S. Chui, "Evaluation of deep inspiration breath-hold lung treatment plans with Monte Carlo dose calculation," Int. J. Radiat. Oncol. Biol. Phys. **53**, 1058–1070 (2002).

[247]R. Timmerman, L. Papiez, R. McGarry, L. Likes, C. DesRosiers, S. Frost, and M. Williams, "Extracranial stereotactic radioablation: results of a phase I study in medically inoperable stage I non-small cell lung cancer," Chest **124**, 1946–1955 (2003).

[248]L. Wang, E. Yorke, and C. S. Chui, "Monte Carlo evaluation of tissue inhomogeneity effects in the treatment of the head and neck," Int. J. Radiat. Oncol. Biol. Phys. **50**, 1339–1349 (2001).

[249]C. Boudreau, E. Heath, J. Seuntjens, O. Ballivy, and W. Parker, "IMRT head and neck treatment planning with a commercially available Monte Carlo based planning system," Phys. Med. Biol. **50**, 879–890 (2005).

[250]J. Seco, E. Adams, M. Bidmead, M. Partridge, and F. Verhaegen, "Head-and-neck IMRT treatments assessed with a Monte Carlo dose calculation engine," Phys. Med. Biol. **50**, 817–830 (2005).

[251]C. Martens, N. Reynaert, C. De Wagter, P. Nilsson, M. Coghe, H. Palmans, H. Thierens, and W. De Neve, "Underdosage of the upper-airway mucosa for small fields as used in intensity-modulated radiation therapy: A comparison between radiochromic film measurements, Monte Carlo simulations, and collapsed cone convolution calculations," Med. Phys. **29**, 1528–1535 (2002).

[252]F. Verhaegen, I. J. Das, and H. Palmans, "Monte Carlo dosimetry study of a 6 MV stereotactic radiosurgery unit," Phys. Med. Biol. **43**, 2755–2768 (1998).

[253]F. Sánchez-Doblado *et al.*, "Ionization chamber dosimetry of small photon fields: A Monte Carlo study on stopping-power ratios for radiosurgery and IMRT beams," Phys. Med. Biol. **48**, 2081–2099 (2003).

[254]A. Chaves, M. C. Lopes, C. C. Alves, C. Oliveira, L. Peralta, P. Rodrigues, and A. Trindade, "A Monte Carlo multiple source model applied to radiosurgery narrow photon beams," Med. Phys. **31**, 2192–2204 (2004).

[255]J. S. Li *et al.*, "Clinical implementation of intensity-modulated tangential beam irradiation for breast cancer," Med. Phys. **31**, 1023–1031 (2004).

[256]J. J. DeMarco, I. J. Chetty, and T. D. Solberg, "A Monte Carlo tutorial and the application for radiotherapy treatment planning," Med. Dosim. **27**, 43–50 (2002).

[257]A. Leal, F. Sanchez-Doblado, R. Arrans, J. Rosello, E. C. Pavon, and J. I. Lagares, "Routine IMRT verification by means of an automated Monte Carlo simulation system," Int. J. Radiat. Oncol. Biol. Phys. **56**, 58–68 (2003).

[258]G. X. Ding, D. M. Duggan, C. W. Coffey, P. Shokrani, and J. E. Cygler, "First macro Monte Carlo based commercial dose calculation module for electron beam treatment planning—New issues for clinical consideration," Phys. Med. Biol. **51**, 2781–2799 (2006).

[259]D. W. O. Rogers, A. F. Bielajew, and A. E. Nahum, "Monte Carlo calculations of electron beams in standard dose planning geometries," in *Proceedings of the 8th ICCR* (IEEE, New York, 1984), pp. 140–144.

[260]C. Scherf, J. Scherer, and L. Bogner, "Verification and application of the Voxel-based Monte Carlo (VMC++) electron dose module of oncentratrade mark MasterPlan," Strahlenther. Onkol. **183**, 81–88 (2007).

[261]K. R. Shortt, C. K. Ross, A. F. Bielajew, and D. W. O. Rogers, "Electron beam dose distributions near standard inhomogeneities," Phys. Med. Biol. **31**, 235–249 (1986).

[262]E. Wieslander and T. Knoos, "A virtual-accelerator-based verification of a Monte Carlo dose calculation algorithm for electron beam treatment planning in clinical situations," Radiother. Oncol. **82**, 208–217 (2007).

[263]J. Cygler, J. J. Battista, J. W. Scrimger, E. Mah, and J. Antolak, "Electron dose distributions in experimental phantoms: a comparison with 2D pencil beam calculations," Phys. Med. Biol. **32**, 1073–1086 (1987).

[264]J. A. Hayman *et al.*, "Dose escalation in non-small-cell lung cancer using three-dimensional conformal radiation therapy: Update of a phase I trial," J. Clin. Oncol. **19**, 127–136 (2001).

[265]I. J. Chetty, M. Rosu, D. L. McShan, B. A. Fraass, and R. K. Ten Haken, "The influence of beam model differences in the comparison of dose calculation algorithms for lung cancer treatment planning," Phys. Med. Biol. **50**, 801–815 (2005).

[266]F. M. Kong, R. K. Ten Haken, M. J. Schipper, M. A. Sullivan, M. Chen, C. Lopez, G. P. Kalemkerian, and J. A. Hayman, "High-dose radiation improved local tumor control and overall survival in patients with inoperable/unresectable non-small-cell lung cancer: Long-term results of a radiation dose escalation study," Int. J. Radiat. Oncol. Biol. Phys. **63**, 324–333 (2005).

[267]I. J. Chetty, "Monte Carlo treatment planning: The influence of 'variance reduction' techniques (ECUT, PCUT, ESTEP) on the accuracy and speed of dose calculations," Med. Phys. **32**, 2018 (abstract) (2005).

[268]www.irs.inms.nrc.ca/inms/irs/papers/iccr00/iccr00.html.

[269]E. Heath, McGill University (personal communication).

[270]E. Poon and F. Verhaegen, McGill University (personal communication).

The British Journal of Radiology, 79 (2006), S2–S15

New developments in MRI for target volume delineation in radiotherapy

[1,2]V S KHOO, FRACR, FRCR, MD and [3]D L JOON, FRACR

[1]Royal Marsden Hospital, Institute of Cancer Research, Fulham Road, London SW3 6JJ, [2]University of Manchester, Manchester, UK and [3]Austin Health Radiation Oncology Centre, Heidelberg Repatriation Hospital, Victoria, Australia

ABSTRACT. MRI is being increasingly used in oncology for staging, assessing tumour response and also for treatment planning in radiotherapy. Both conformal and intensity-modulated radiotherapy requires improved means of defining target volumes for treatment planning in order to achieve its intended benefits. MRI can add to the radiotherapy treatment planning (RTP) process by providing excellent and improved characterization of soft tissues compared with CT. Together with its multiplanar capability and increased imaging functionality, these advantages for target volume delineation outweigh its drawbacks of lacking electron density information and potential image distortion. Efficient MR distortion assessment and correction algorithms together with image co-registration and fusion programs can overcome these limitations and permit its use for RTP. MRI developments using new contrast media, such as ultrasmall superparamagnetic iron oxide particles for abnormal lymph node identification, techniques such as dynamic contrast enhanced MRI and diffusion MRI to better characterize tissue and tumour regions as well as ultrafast volumetric or cine MR sequences to define temporal patterns of target and organ at risk deformity and variations in spatial location have all increased the scope and utility of MRI for RTP. Information from these MR developments may permit treatment individualization, strategies of dose escalation and image-guided radiotherapy. These developments will be reviewed to assess their current and potential use for RTP and precision high dose radiotherapy.

Received 9 August 2005
Revised 6 January 2006
Accepted 10 March 2006

DOI: 10.1259/bjr/41321492

The increased sophistication of modern radiotherapy planning techniques such as conformal (CFRT) and intensity-modulated radiotherapy (IMRT) necessitates improved means of defining target volumes for treatment. This is needed to achieve the intended benefits of using CFRT and IMRT. This step remains the most crucial and difficult part of the radiotherapy planning process, otherwise a geographical miss of the tumour or a systematic error will be perpetuated throughout therapy. MRI is being increasingly used in oncology for staging, assessing tumour response and evaluating disease recurrence. As a result of the enhanced imaging properties of MR, it has been estimated to be a more cost effective diagnostic tool in the management of some diseases [1]. Similarly, the improved characterization of soft tissues and visualization of tumour extent using MRI can be used to benefit the radiotherapy treatment planning (RTP) process from delineation of target volumes to determining planning margins and treatment response [2].

There are many current areas of development in MRI. These include developments in hardware technology, such as 3 Tesla machines, and the use of new MR contrast media, such as ultrasmall superparamagnetic iron oxide particles for lymph node evaluation [3]. MR techniques and sequences previously used for research are now becoming available for general use. MR techniques such as dynamic contrast enhanced and diffusion weighted MRI may provide further characterization of tissue and tumour regions [4, 5]. MR sequences such as ultrafast volumetric and 3D cine sequences can offer the opportunity to assess target/organ motion and deformity [6, 7]. Temporal-spatial information gleaned from MRI can then be used for image-guided strategies in radiotherapy delivery. All these features have the potential to increase the scope and utility of MRI for RTP.

It is worthwhile briefly reviewing the background to the use of MRI for RTP in order to understand the rationale and issues with its use. Some methods of utilizing MRI in RTP will be outlined. This article will then discuss the new MRI developments in terms of their current and potential impact in target volume definition for treatment planning with examples of applications at some cancer subsites.

MRI rationale for RTP

Any additional procedures used for RTP must add value to the planning process. Standard RTP uses CT data. CT images are good at distinguishing between structures that have substantially different X-ray attenuation properties or Hounsfield units, such as between air, tissue and bone. It is more difficult to discriminate between adjoining soft tissue structures using CT if these soft tissue structures possess similar

Hounsfield units unless there is a fat, air or bone interface between these structures. The imaging parameters for CT scanning are much more limited compared with the range available with MRI.

In the case of MRI, the contrast from soft tissue structures can be widely varied by extensively manipulating the imaging parameters, which include proton densities and tissue relaxation times (spin-lattice or T_1 and spin–spin or T_2). This increased flexibility in varying tissue contrast or signal intensities offers much better characterization of soft tissues even when these structures possess very similar X-ray attenuation properties or electron densities. Tumours often have similar electron densities to their neighbouring soft tissues. By using different MRI sequences, better tissue discrimination can be obtained between the extent of tumour with its boundaries of infiltration and the adjacent normal structures. In this manner, MRI provides improved target delineation for RTP. This utility of MRI applies not only to the initial radiotherapy treatment of tumours but also potentially for re-treatments by being able to differentiate between changes due to recurrent cancer or that secondary to post-treatment fibrosis. It can also provide better delineation of organs at risk (OARs) for dose avoidance in RTP.

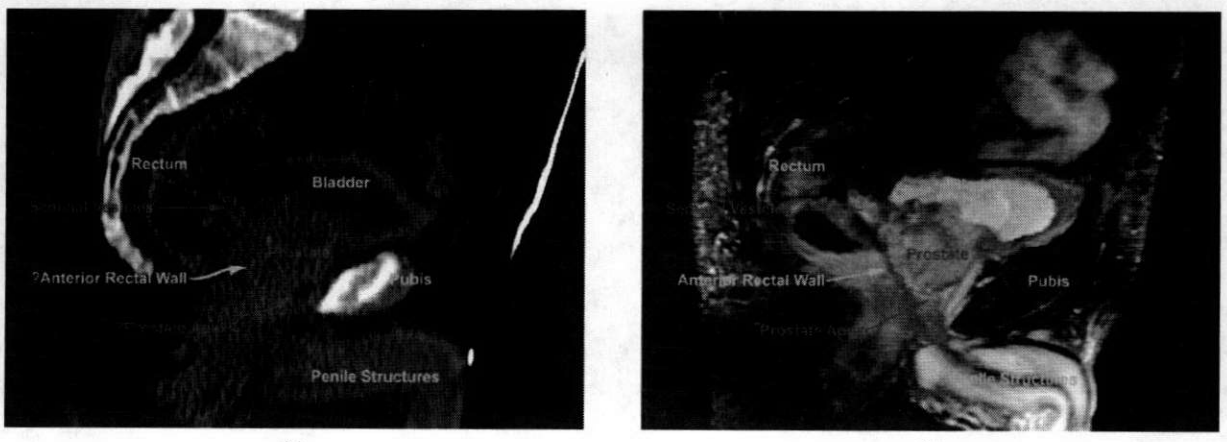
An obvious benefit of enhancing the visualization of volumes of interest (VOI) is the increased reliability and consistency of target definition. This will improve both interobserver and intraobserver variability for outlining. This has value for institutional and multicentre trials in radiotherapy where it is important to maintain consistent and accurate target and OAR volumes. Substantial and inappropriate variations in target volumes can impact on trial outcomes, with geographical misses for poorer local control rates or unnecessary inclusion of normal tissue for higher toxicity rates.

MRI can avoid bony and metal artefacts seen with CT. Large thick bony sections attenuate X-rays and reduce the adjacent soft tissue image quality. This can obscure identification of nearby tumours and internal anatomy. MR images are not affected by this. MRI can thus further improve the delineation of both tumour and OAR volumes for RTP in these regions.

Another feature of the increased functionality of MRI is its true multiplanar capability. This ability to image in any oblique plane can reduce the "partial volume" imaging effect that often results from conventional transaxial CT imaging, particularly where the 3D shape of the target is extreme or changes substantially between conventional transaxial CT slices. Para-sagittal or para-coronal views can also permit better understanding of the boundaries of target volumes with the surrounding normal tissues leading to better target volume delineation (Figures 1 and 2).

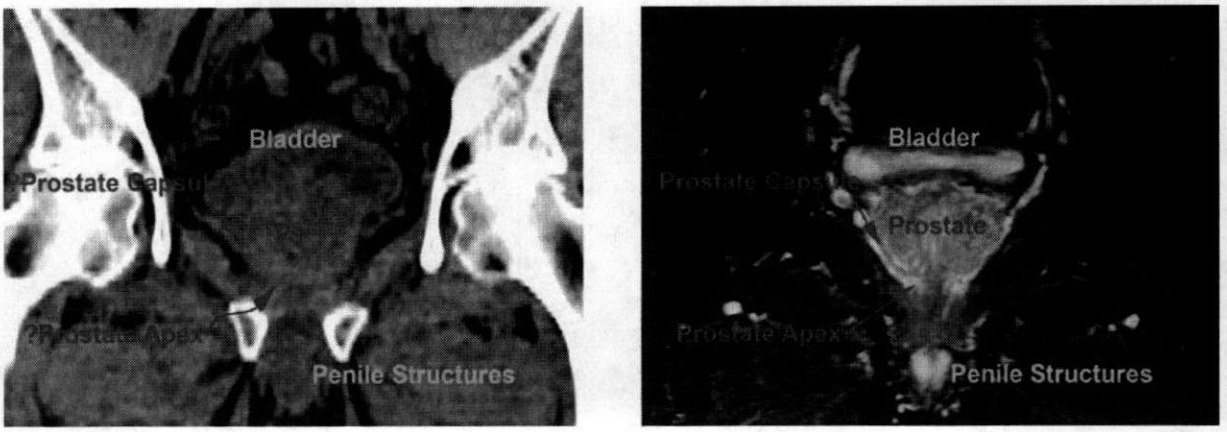
Furthermore, MRI can provide functional and biological information for tumour regions that may improve target definition and permit new opportunities for novel radiotherapy strategies. Some of the salient features of using MRI for RTP are summarized in Table 1. However, it is important to be aware that implicit in the use of MRI for RTP is that oncology clinicians should have the necessary training to comprehend MR images and understand how to use them appropriately for defining VOIs. Thus it is important to undertake suitable supervised training [8]. Even if there is relevant experience, it is still beneficial to liaise closely with local diagnostic colleagues who have MR expertise for the cancer subsite(s) in question. Ideally there should be an oncology team collaboration for the definition of target volumes in radiotherapy similar to the multidisciplinary team arrangement that exists for general cancer management. This approach has been endorsed by national bodies, such as the Royal College of Radiologists in a recent publication [9].

All of these advantages add to the RTP process and outweigh its current drawbacks, which include the lack of electron density information, potential image distortion and specific patient considerations with MR scanners. Some examples of the current impact of using MRI for target volume delineation and some of these



(a) (b)

Figure 1. A comparison of sagittal views of the pelvis for prostate radiotherapy with (a) CT reconstructed from 2.5 mm slices and (b) MR image obtained in-plane in the same patient. Some of the relevant structures of interest for radiotherapy are labelled on the MR image. These structures are not visualized well enough on CT to provide confident determination of the prostate boundaries for radiotherapy.



(a) (b)

Figure 2. A comparison of coronal views of the pelvis for prostate radiotherapy with (a) CT reconstructed from 2.5 mm slices and (b) MR image obtained in-plane in the same patient. Definition of the prostate gland boundaries and the adjacent structures is better visualized on MRI than with CT.

Table 1. Advantages and disadvantages of MRI for radiotherapy planning (RTP)

Features	Advantages	Disadvantages
Patient	Non or minimally invasive procedure	Claustrophobia due to the smaller patient bore
	Few patient risks	Contraindicated in patients with loose metal foreign bodies within the body, particularly the orbits or pacemakers
	No radiation associated with imaging	
	This may be advantageous to paediatric patients and pregnant women	
	This may be a useful for follow-up scanning	
Imaging	Increased number of imaging parameters for more imaging flexibility	MR image distortion
	Superior soft tissue imaging with excellent spatial resolution to provide better visualization for the following: determining the tumour/GTV extent and degree of tumour infiltration	Systems
	Understanding the surgical bed or altered anatomy secondary to surgery	Object induced distortions
	Distinguishing between post-treatment fibrosis or tumour recurrence	Lack of electron density information for dosimetry and needs additional steps to permit dose calculations
	Improved definition of normal soft tissue structures and tissue planes avoidance of image artefact from metal prosthesis and large bony regions	Lack of cortical bone information to create digitally reconstructed radiographs (DRR) in radiotherapy
	True multiplanar capability to image in any oblique plane and reduction of the "partial volume" imaging effect increased accuracy, reliability and consistency of target definition to reduce both interobserver and intraobserver variability	May have longer scan times than CT with more potential for motion artefacts
	Providing functional and biological information for functional avoidance or biological targeting	Need for specific training to comprehend and understand MR images for RTP use
	Ultra-fast volumetric and cine mode acquisitions to assess temporal-spatial variations in target positioning or deformation	RTP systems can only import transverse MR images and cannot take full advantage of sagittal and coronal in-plane MR images
	Can be registered with CT information for use in RTP systems	Immobilization devices used in radiotherapy may not be MR compatible
Contrast agents	New contrast agents (*i.e.* USPIO) to define nodal status	
	Less incidence of allergic reactions to gadolinium than iodine-based contrast agents	
Machine	New bore flange openings to reduce patient claustrophobia	Not as readily available and accessible as CT
	Open MR systems for easier patient access, tolerance and positioning for radiotherapy	Smaller bore than CT (52 cm *vs* 82–85 cm)
		Curved table top

GTV, gross tumour volume; USPIO, ultrasmall paramagnetic iron oxide.

issues in using MRI for RTP will be briefly outlined below. A more in-depth review has been published and the reader is advised to peruse this [10].

Current impact of MRI for treatment planning

In many areas of oncology, diagnostic MRI is the gold standard imaging modality for the staging and assessment of cancer patients. This is true for imaging of cerebral and spinal tissues, soft tissue sarcomas and pelvic tumours. Some examples will be highlighted here to illustrate the current status of the added advantage of MRI for target volume delineation in the planning process.

MRI has been used extensively for central nervous system (CNS) radiotherapy. Many investigators have reported quantitative improvements of up to 80% of cases in target volume definition with the addition of MRI to 3D CT based treatment planning [11–13]. Despite the obvious advantage of MRI for target volume definition at this site, there are clear situations where the use of both CT and MRI data is valuable and can provide for more consistent volume delineation than compared with either MRI or CT alone [14]. In a study of base of skull meningiomas, MRI was able to delineate tumour volumes that were present close to the base of skull bones as the X-ray attenuation from these large bones can obscure soft tissue detail using CT alone [15]. However, CT was able to provide information on the extent of bony erosion from tumour that was not available with MRI. In some cases, the individual CT or MR volumes were vastly different, with each modality providing separate but complementary information (Figure 3). This supports the use of combined CT and MRI data to provide the optimum target volume delineation. Segmentation algorithms, automated and atlas based, may further aid the delineation process and this methodology is currently being investigated [16, 17]. In many centres providing CFRT and IMRT for CNS tumours, the use of CT-MRI fusion for RTP may now be considered as standard practice. The major thrust of current developments in CNS treatment planning is to integrate the use of functional data for target volume determination [18].

Head and neck anatomy is complex and the extent of the infiltrating tumour can be difficult to define. MRI can assist in delineation of radiotherapy volumes here. It is useful for defining (1) longitudinal tumour infiltration along the upper aero-digestive tract and adjacent fascial planes, *e.g.* pre-vertebral fascia, (2) tumour infiltration of soft tissue structures and tissue planes such as the pterygoids and tongue (Figure 4), (3) the extent of perineural infiltration and intracranial extension, *e.g.* nasopharyngeal tumours and (4) nodal metastases. This situation is best illustrated by nasopharyngeal tumours where the use of multimodality imaging with MRI can change disease staging in about 50% of cases impacting on RTP [19, 20]. In a study of over 250 patients, up to 40% of intracranial infiltration detected on MRI was missed by CT [21]. Segmentation algorithms for MRI are being developed to help in target volume delineation based on the contrast enhancement ratio of T_1 weighted images and signal intensity of T_2 weighted images, and this methodology may further assist the RTP process [22].

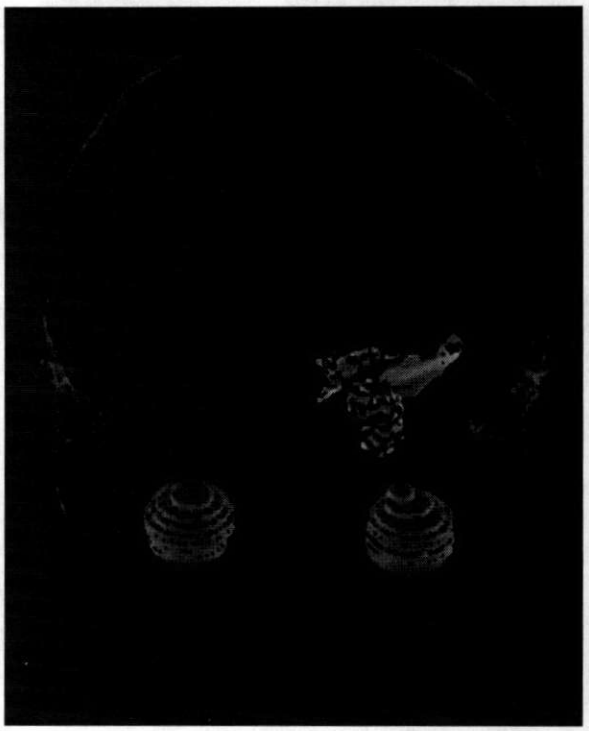
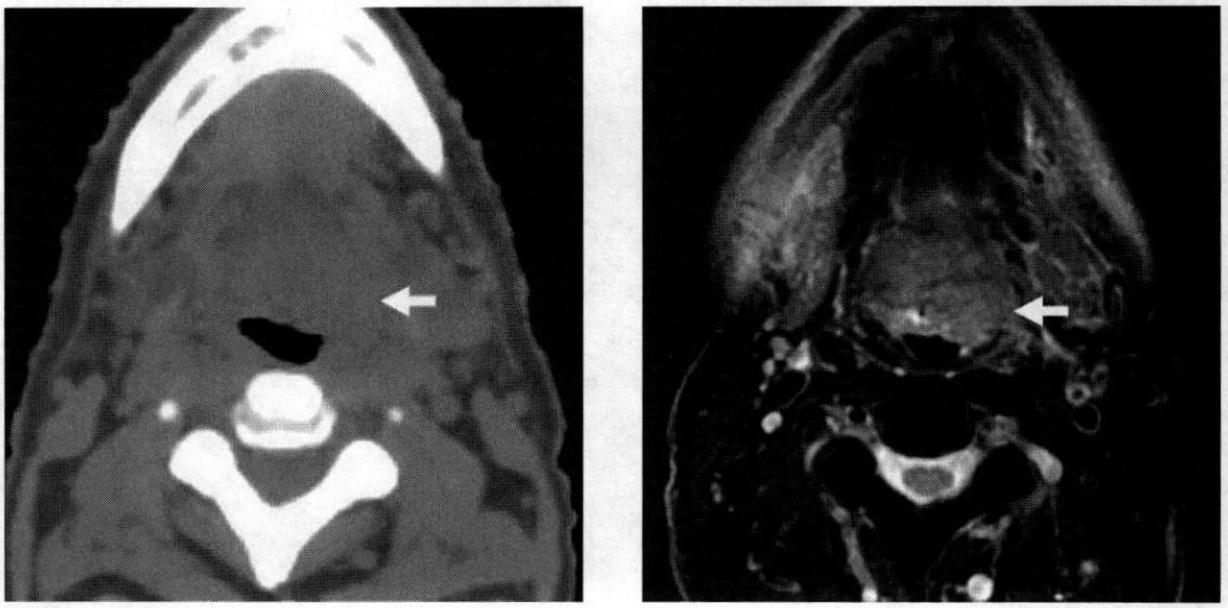


Figure 3. A case from a study of meningiomas of the skull base evaluating the differences between MRI and CT assessment of the clinical target volume (CTV) for radiotherapy. The 3D reconstructed view of the CT-defined CTV (red outlines) and MR-defined CTV (yellow outlines) illustrates the spatial differences in CTV definition by the two different imaging modalities where the MR-defined CTV demonstrates tumour extending laterally along the petrous ridge that was not seen using CT [15].

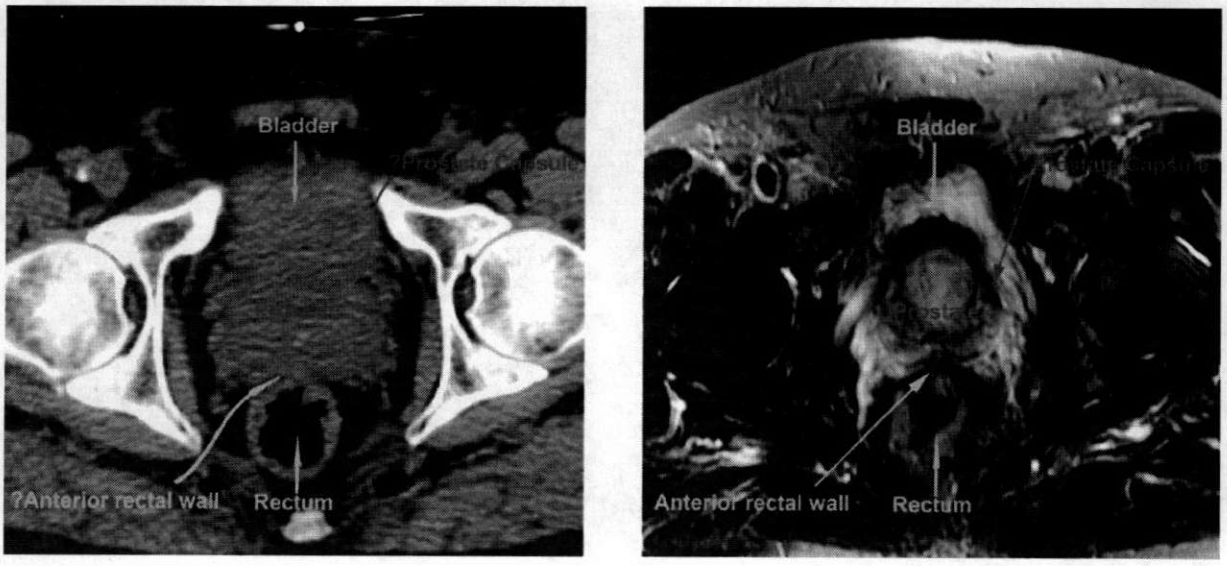
In the pelvis, MRI has provided improved target delineation for urological, gynaecological and gastrointestinal cancers. For prostate cancer, MRI can provide better internal organ assessment than CT for disease extent, capsular and seminal vesicle involvement [23–25]. MRI can overcome some of the limitations of CT definition of prostate treatment volumes (Figures 1 and 2). MRI, particularly using sagittal views, can be useful in defining the prostatic apex and distinguishing between the boundaries of the prostate with the base of the bladder and the anterior wall of the rectum [26, 27]. The prostatic capsule, which cannot be distinguished from adjacent normal tissue on CT, can be seen as a thin rim of low signal intensity on T_2 weighted MRI thereby permitting boundary definition (Figure 5). Comparative MRI-CT planning studies using MRI-defined prostate volumes as the gold standard have reported that CT-defined prostate volumes tend to overestimate the planning volume by as much as 27–43% due to the soft tissue uncertainty in CT delineation [28–31]. In our prospective study of 105 men with prostate cancer, we found that MRI-CT fusion for RTP can often up-stage the disease extent by clarifying extracapsular spread, seminal vesicle involvement and early adjacent organ invasion that resulted in substantial changes to the target



(a) (b)

Figure 4. A case of a base of tongue cancer imaged with (a) CT and (b) MRI showing a large mass on the left side of the oropharynx involving the base of tongue and the left tonsillar fossa with invasion of the left parapharyngeal space. The base of tongue mass extends past the midline. These features are better visualized using MRI than CT.

volumes in 29% of cases [32]. In a retrospective review of 199 patients treated with radiotherapy, a comparable upstaging effect of MRI was seen in 52% [33]. For post-prostatectomy patients, a visible mass was noted within the operative bed using MRI in 50% of cases, that was not seen using CT [32]. This improved ability to delineate prostate and seminal vesicles can also reduce inter-observer and intraobserver variation. MRI can also be useful where the internal pelvic anatomy has been substantially altered due to previous extensive surgery such as abdominal-perineal resections [34]. It can also aid delineation of adjacent normal tissue structures such as rectal wall, recto-vesicle fascia of Denonvillier, urogenital diaphragm, penile bulb, periprostatic venous plexus, neurovascular bundle, levator ani and anal sphincters. The use of MR-based prostate planning volumes can



(a) (b)

Figure 5. A comparison of transaxial co-registered views of the pelvis for prostate radiotherapy with (a) CT and (b) MR in the same patient. The boundaries of the prostate are better visualized using MR than with CT, notably the anterior rectal wall/recto-vesical fasia and prostate capsule.

result in more appropriate treatment volumes, leading to better shaping of the treatment fields that can reduce the risk of treatment related complications to important normal structures such as the rectum and penile bulb [35].

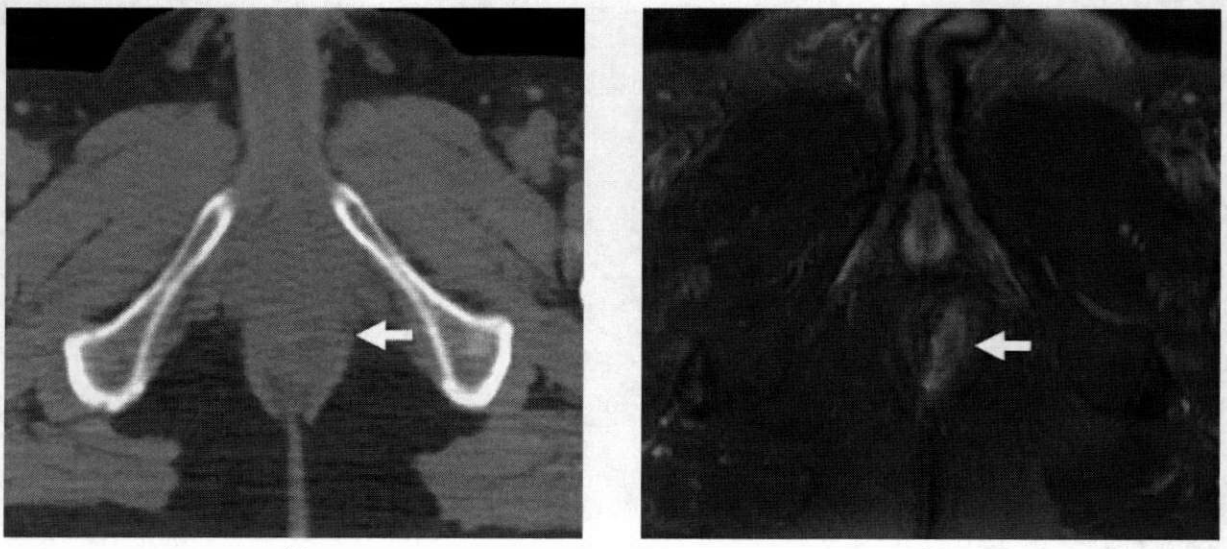
In addition, co-registered CT/MRI has also been used for permanent and high dose rate prostate brachytherapy to better define target volumes, reduce delineation uncertainty for RTP and to guide needle insertion compared with other imaging methods used such as ultrasound or CT [36, 37]. This is particularly important for post-implant dosimetry where the prostate gland delineation may be more difficult following the implant procedure and where CT assessments may be hampered by its poor intrinsic tissue contrast and seed induced artefact [38, 39]. MR-CT fusion has also been found to be beneficial for brachytherapy in other tumour sites such as head and neck, sarcomas and gynaecology [40]. In cervical brachytherapy, the use of MRI has been strongly supported by the Gynaecological GEC-ESTRO Working Group [41].

For rectal cancer radiotherapy, traditional planning relies on visualizing a filling defect using rectal contrast with the treatment fields usually placed according to bony landmarks, but rectal contrast does not define the circumferential thickness of the tumour and bony landmarks do not accurately define the anterior margin as governed by the rectal lymphatic drainage and mesorectum. CT planning can improve delineation of the rectal tumour by providing visualization of the increased thickness of the rectal wall and provide better definition of the lymphatic region by depicting the vascular structures, visible lymph nodes and boundaries of the mesorectum [42]. However, inaccuracies in the CT definition of tumour may occur because of poor contrast between faeces and tumour, partial volume effects due to the curves/valves of Houston in the rectum and imaging of the horizontal sigmoid. Infiltration into the anus can also be difficult to assess in a low rectal tumour unless there is an obvious mass effect (Figure 6). MRI can avoid these CT identification issues and better define the depth of invasion through the rectal wall [43]. MRI can provide visualization similar to endorectal ultrasound investigations, but ultrasound data cannot be imported into treatment planning systems. MRI can aid the CT based RTP process by defining the longitudinal spread of the cancer superiorly and inferiorly, as well as the extent of infiltration of the mesorectum [44]. It may also provide better assessment of early invasion into local adjacent structures such the bladder, prostate or seminal vesicles in men, and vagina and uterus in women, in addition to anal sphincter infiltration in both [45–47]. This better estimation of the gross tumour volume may permit strategies of anal sphincter sparing, tumour boosting and dose escalation with or without concurrent chemotherapy.

MRI issues and schemas for treatment planning

A comprehensive step-by-step guide and methodology review for the integration of MR into treatment planning systems is outside the remit of this article. However, it is worthwhile to outline briefly some of the relevant issues faced when using MR images for RTP and some of the available methods used.

In RTP, geometrically accurate images are needed for precision radiotherapy and electron density information is required to take into account tissue inhomogeneities when calculating dose distributions. CT data provide the necessary information for both these requirements. MRI suffers from a lack of electron density information and potential spatial image distortion. As a result, MR images cannot be imported alone into RTP systems to readily create and plan three-dimensional (3D) radiotherapy.



(a) (b)

Figure 6. Co-registered (a) CT and (b) MRI scans showing a rectal cancer in the lower rectum extending to the anorectal junction with invasion of the left posterolateral wall. This is easily seen on the T_2 weighted MRI scan as a hyperintense signal region that is not visible on CT.

Unlike CT images where the Hounsfield units can be directly related to electron density values, the signal intensities in MR images do not have such a relationship. If MR images are to be used alone for RTP, then X-ray attenuation coefficients have to be manually assigned to all regions within the MR image in order to allow dosimetry calculations. In some regions, such as in the brain where the internal tissue is more homogeneous and the skull can be easily segmented, this may be easily achieved by assuming an homogeneous attenuation value within the skull vault. On MR images, investigators have calculated that doses were within 2% of those calculated using CT data [48]. Similar work has been recently performed in the pelvis for prostate radiotherapy with 2–3% between CT-based and MR-based dose calculations [49]. Other investigators noted dosimetry discrepancies that were >2% compared with CT dosimetry for the pelvis if bone and water CT number bulk-assigned values were not assigned to MR images [50]. Simple methods of CT density value assignments may more difficult to implement if the anatomy is varied or complex such as in the head and neck regions or thorax/abdominal regions.

An alternative method is to integrate MR images with CT data. This requires bringing the MR data into spatial 3D alignment with planning CT data or image registration. Then the different image modalities can be "fused" so that a common image set is produced with both sets of imaging information available for clinical interrogation. This method is favoured as the data can be transferred and implemented within a treatment planning system and provides for optimum evaluation of the imaging information to create the most appropriate VOI for RTP. The methodology for image co-registration and image fusion is considered by Kessler in this issue.

MR image distortion

Before image co-registration can occur, it is important to ensure that the MR data are suitable. MR image distortion is one potential concern. MR image distortions can be grouped into two main categories; system-related and object-induced. An illustration of these distortion effects is shown in Figure 7.

System-related distortions are due to the imperfections of the magnet, its operating system and imaging sequences. The effects of magnet field inhomogeneity, gradient field non-linearities and eddy currents can contribute to system-related image distortion. In general, the effects of system related distortion are smallest at the centre of the magnet and worsen with increasing distance from the magnet centre. The distortion magnitude is largest at the periphery of the field-of-view (FOV).

Object-induced distortion arises when any object (*i.e.* the patient) is placed within a magnetic field. This type of distortion results from magnetic susceptibility and chemical shift effects. Different body tissues have different magnetic susceptibilities and susceptibility artefacts can be pronounced at different tissue boundaries, such as between air cavities and soft tissues. Chemical shift effects result from the different behaviour of protons in fat and tissue. Fat protons "precess" at a slower rate than water protons and this can result in a chemical shift effect where the positions of fat/water protons are shifted from their true spatial locations.

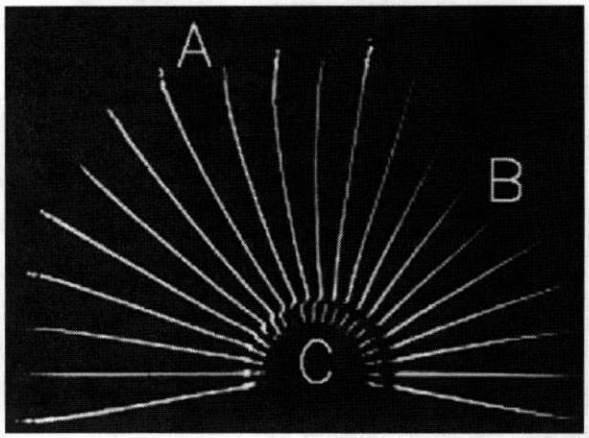


Figure 7. An illustration of various forms of distortion in MRI using a phantom consisting of a coplanar array of water-filled tubes embedded within in a circular solid plastic (PMMA) block. System distortion effects are seen in the apparent curvature of the tubes at A and their disappearance at B, which was due to warping distortion of the imaging plane. Magnetic susceptibility differences due to the presence of the plastic support block at C give rise to object-induced distortions in the form of discontinuities at the point where each tube enters the support block [10]. (Reproduced with permission from Elsevier).

In order to utilize MR images for RTP, these image distortions must be evaluated, minimized, and/or corrected especially if they are to be co-registered with CT, otherwise a systematic error will be incorporated into the treatment plan. There are many methods to deal with MR image distortion, but the primary task is to quantify the presence and extent of any distortion. This methodology has been previously described in detail [51–54]. This system provides for evaluation and correction of both system-related and object-induced distortions, but can also facilitate quality assurance programs. It is important to ensure that the same MR scanner and imaging sequence used for mapping is also used for patient imaging.

System-related distortion is best quantified and mapped using a phantom (linearity test object) with a known array of markers in 3D to provide spatial assessments of the whole imaging volume used for RTP. A reference frame with imbedded markers provides another set of marker positions that covers the periphery of the imaging volume. The linearity test object or the patient is housed within this reference frame. A separate set of markers placed on the patient's surface provides assessment of object-induced effects. In brief, the positions of all markers are mapped within the imaging volume using a dedicated automated algorithm. The use of read-out gradient reversal imaging and post-processing image corrections can account for object-induced effects. Distortional shifts of up to 5 mm can be corrected [51–54]. Other correction methods have also been developed and can reduce distortions by a factor of two [55]. Site specific phantoms have also been created

with air cavities to assess the use of MRI for lung RTP [56].

Other MR considerations

It is important to review the image protocols to be used and to evaluate those sequences which can offer the best combination of image quality/resolution and minimal image distortion. These sequences will differ depending on the anatomical site being imaged and treated. As previously mentioned, the effect of distortion is least within the centre of the magnet and VOI for RTP should be imaged within this zone whenever possible. If larger FOVs are used, the central region of imaging may be extracted and this region can then be co-registered with CT data to aid target volume definition. In determining imaging protocols for multi-modality image co-registration, attention must be paid to some of the issues listed below to optimize the appropriate imaging acquisition parameters.

1. What is the appropriate sequence to be used for the cancer type and anatomical region? For example, will it be T_1, T_2 weighted or a hybrid sequence, standard or ultra-fast acquisition, gated or volumetric, and contrast or not? This will depend on what imaging information is needed by the clinician.
2. What is the optimal FOV and relevant imaging volume?
3. What is the resolution needed?
4. What is the appropriate imaging slice orientation, slice thickness and slice gaps?
5. Is there a need for multislice and/or non-coplanar image reconstruction or oblique plane imaging?
6. What is the influence of various body coils or internal body MR probes on the MR images for co-registration?
7. What is the appropriate quality assurance program to ensure reliability of image quality and data transfer?

Any MRI for RTP should also mimic the CT planning procedures such as scanning the patient in the treatment position with a flat bed insert, using the same but MR compatible immobilization devices where specified, providing the same instructions to the patient, *e.g.* full or empty bladder, and minimizing internal organ motion by breath held procedures, bowel relaxants or reducing the scanning time whenever indicated and possible. Ideally the MRI scans should be timed as close as possible to the CT planning scans.

New developments in MRI

Developments in MR hardware

3 T MRI scanners

The field strengths of MRI scanners have been increasing since their initial development from 0.5 T (Tesla) scanners to 1 T and now 1.5 T and 3 T. The advantages of increasing the field strength are several-fold. The quality of the MR image is related to the signal obtained relative to any imaging deficiencies using the MR scanner, such as machine imperfections and object/patient induced errors. This relationship is termed the signal-to-noise ratio (SNR). The image quality or resolution increases when the signal is strengthened and the noise lowered. The SNR approximately increases in a linear manner with field strength. In simplistic terms, higher field strengths would permit higher resolution images with improved tissue contrast that can lead to better tissue diagnosis and definition of tumour volumes for RTP. The image acquisition time may also be reduced and this can potentially minimize motion artefacts. With these advantages, the development and use of 3 T MRI machines may become more commonplace as they can further improve the image quality using external phased array coils. In prostate MRI, the image quality from external phased array coils using 3 T are comparable with endorectal coils at 1.5 T [57] and in some cases exceeding them [58]. This provides more imaging options and avoids the internal deformation that occurs with endorectal coils that may limit its use for RTP. These 3 T machines can also benefit the application of functional MRI and MR spectroscopy (MRS) by providing better resolution for assessed metabolites that can define tumour regions, *e.g.* for brain [59] and prostate radiotherapy [60]. The latter subject is covered in more detail by Payne and Leach in this issue.

There are several issues associated with 3 T MR scanners that may limit their use for RTP. Amongst these is the exacerbation of magnetic susceptibility effects, doubling of the chemical shift effect, patient safety and engineering challenges. Currently, the clinical role of these 3 T scanners including their utility for RTP remains to be defined.

Magnetic susceptibility effects. Higher field strengths can exacerbate the magnetic susceptibility artefacts, particularly at borders of different structures; for example, at tissue/bone/air boundaries, such as in the head and neck regions [61]. These effects can lead to increased signal intensities that can cause misdiagnosis using contrast studies unless pre- and post-contrast views are used for appropriate comparison. This effect can potentially result in misidentification of abnormal areas leading to erroneous volumes for RTP.

Chemical shift effect. The chemical shift effect (see above) is doubled using 3 T field strengths. This effect causes misregistration of fat and water tissues and will hinder the definition of radiotherapy volumes when co-registered with planning CT images. However, this aspect may be of considerable use in MRS as it can increase the resolution of identifying tissue metabolites. This may permit improved and more reliable localization of tumour regions for boosting, such as the use of MRS citrate-choline assessment in prostate cancer. It is also very useful for brain and MR angiography assessments because of the longer, more variable T_1 relaxation time.

Patient safety. There may be patient safety concerns as the energy deposited using 3 T scanners may be up to 4 times that of 1.5 machines. This may be a factor when using fast or intensive pulse sequences. Potential

problems may be reduced by appropriately modifying the pulse sequences and/or reducing the volume of tissue being imaged.

Machine factors. 3 T scanners are still not commonplace and are substantially more expensive than 1.5 T machines. There are also engineering issues that include the use of higher gradient systems and increased shielding, as well as smaller scanner bores. These may restrict the size of the patient scanned and also increase the potential for patient claustrophobia.

MR simulators

Open and low field MR scanners can act as radiotherapy simulators providing a radiotherapy environment similar to CT simulation for RTP. Other advantages include the lower cost of these machines, greater T_1 image contrast, reduced vessel flow/ghosting artefacts and the opportunity to use patient immobilization devices that were previously restricted by size with conventional MR scanners. Although open low field MR scanners may have larger system-related distortions which can be compensated for, object-induced distortions are substantially less due to the lower magnetic field strength [40, 62]. Individual MR sequences may be longer using low field MRI and provide more opportunity for internal organ motion, but appropriate selection of sequences can reduce the scanning required. One limitation of MR simulation is that there is inadequate image detail of bony structures from MR to automatically create digitally reconstructed radiographs needed for treatment verification. Software to delineate bone regions can overcome this issue.

The lower SNR may limit diagnostic quality images, but good quality images have been obtained using low field 0.2 T MRI in regions of interest such as the prostate [63]. A recent study of 243 patients revealed that open low field MR simulation provided adequate images for RTP in up to 95% of cases [40]. The greater target volume delineation from MRI led to RTP improvements in up to 33% of lung cases and 40% of prostate cases. MR simulation can better delineate erectile soft tissues to permit dose sparing of these structures by IMRT [64].

Developments in MR sequences

Manipulation of the relaxation times of protons in tissues provides the two basic T_1 and T_2 weighted MR sequences. Whilst this provides superior imaging of soft tissues, the imaging parameters can also be manipulated to benefit RTP by providing ultrafast imaging, volumetric sequences and cine-mode acquisition. These sequences can be used to provide information on target motion, OAR displacement to modify planning margins for RTP and to initiate image-guided radiotherapy strategies.

A series of ultrafast imaging can be obtained by sequences such as echoplanar imaging. Evaluation of the position of OARs and tumours that are influenced by respiration or other internal organ activity during irradiation can be performed and modelled. MRI can provide better delineation of internal soft tissue structures such as between myocardium and ventricular space compared with CT for heart dose–volume assessment by breath hold for left breast radiotherapy [65] or ultrafast acquisitions can be performed with normal respiration to assess the impact of irradiation compared with other treatment image-guided strategies. Ultrafast acquisitions may also be used to obtain gated images to model organ and/or target motion for RTP [6].

Sequential volumetric and cine-mode acquisitions can greatly aid image-guided radiotherapy strategies by providing data for the implementation of appropriate site-specific planning margins for both the target and OARs. Cine MRI can evaluate intrathoracic tumour mobility for patient individualization of treatment margins [66] and determination of the efficacy of free-breathing gating techniques for lung radiotherapy [67]. Cine MR has been used to assess intrafraction motion in prostate cancer [7, 68]. This intrafraction information can be used not only to determine internal margin size for RTP but also to estimate the degree of organ deformation that may occur during radiotherapy [69]. This issue of target volume deformation is also currently being investigated for bladder cancers in an image-guided program (*POLO* or *P*redictive *O*rgan *LO*calization) whereby cine MR studies are obtained to assess the temporal-spatial changes of the bladder as it fills during radiotherapy and the degree of tumour deformation that can occur during this period [70].

Ultrafast, volumetric and cine MRI can provide non-invasive means to evaluate not only variability in target volume positioning during radiotherapy but also the temporal variation in target volume deformation that may occur interfractionally and intrafractionally. These are pertinent issues that currently limit precision radiotherapy and justify the development of image-guided radiotherapy.

Developments in MR contrast agents

Local-regional control is an important issue in radiotherapy treatment for many cancers. Adequate dose to involved lymph nodes can increase the probability of local control and this may translate into improved survival. Reducing dose appropriately to uninvolved local-regional nodal regions can substantially lower the probability of radiation related side-effects and also allow combination with systemic chemotherapy or radiosensitizers to be better tolerated or permit dose intensification. However, the determination of pathological lymph nodes is poor using current imaging techniques and surgery remains the gold standard in establishing the lymph node status.

The difficulty in using simple size criteria alone to assess evidence of subclinical tumour involvement is well known. Up to 20% of normal size lymph nodes can be positive for microscopic disease whilst up to 30% of enlarged lymph nodes may only show inflammation. The likelihood of pathological nodal involvement increases with higher stage of disease. Furthermore, in MRI the signal intensity and degree of contrast enhancement between benign and malignant lymph nodes is not reliable enough to provide satisfactory discrimination.

Ultrasmall superparamagnetic iron oxide particles (USPIO)

USPIO particles have recently been reported to be suitable as contrast agents for the identification of and discrimination between normal and abnormal lymph nodes.

USPIO particles are injected and taken up by macrophages and transported via the lymphatic system to the lymph nodes. In the reticuloendothelial tissues of normal lymph nodes, these macrophage ingested USPIO particles produce a reduction in signal intensity within the node due to the negative enhancement from the iron oxide particles. This lowering of signal intensity when scanned 24–26 h following administration can then be compared with the pre-USPIO scan. An example of this negative enhancement seen in normal lymph nodes is shown in Figure 8. In diseased nodes that are replaced by tumour, these USPIO particles within macrophages are prevented from occupying the node by the tumour and thus the signal intensity from pathologically involved nodes is preserved.

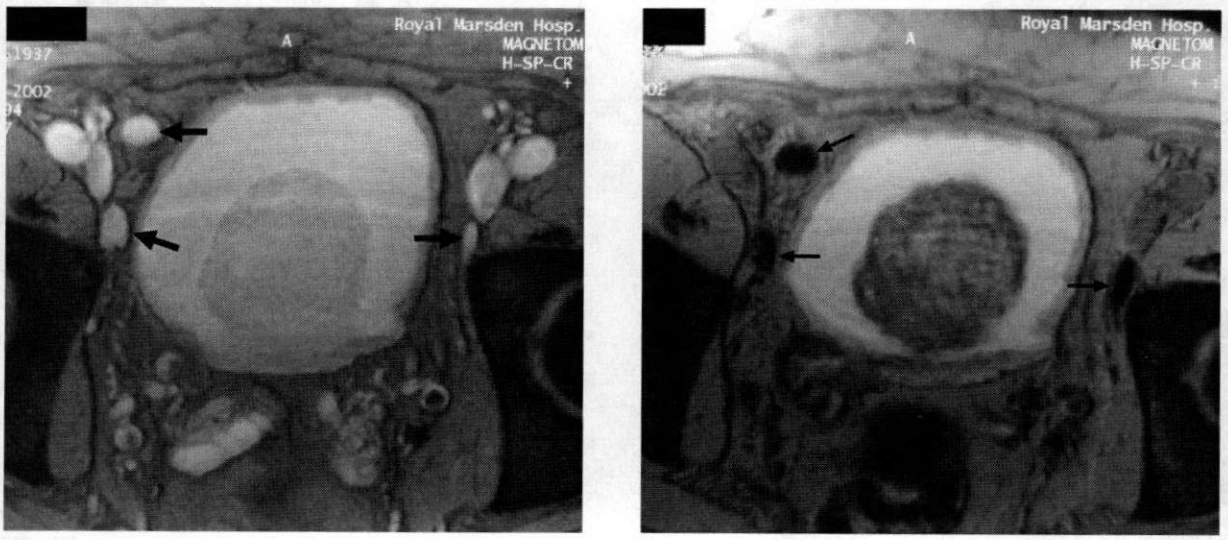
Recent assessments using USPIO have reported a high sensitivity and specificity of up to 90% in cases with small volume tumour nodal involvement [3, 71]. If future studies confirm that this method of detecting early nodal involvement is reliable then the benefits for RTP are immense. This method using 3D sequences can provide visualization of lymph nodes along blood vessels and allow for mapping of lymph nodes according to surgical templates. This will optimize current target volume delineation for lymph nodes. There are many tumour regions where CT defined nodal volumes have only just superseded traditionally designed nodal volumes that are based on regional and bony landmarks. Inadequate nodal coverage can either compromise local control or unnecessarily irradiate normal structures [72, 73].

The use of USPIO imaging can permit individualization of treatment fields. It can substantially influence local-regional volumes currently being prophylactically treated or permit dose escalation for involved lymph nodes. The treatment of head and neck tumours exemplifies this management challenge. USPIO imaging has been used for head and neck planning of surgery [71]. It can also complement the current recommendations for CT defined nodal volumes for head and neck radiotherapy [74]. It is important to note that whilst USPIO methods provide an advance in identification of pathological lymph nodes, it remains a morphological method and hence has its limitations. The threshold size for detecting pathological involvement may be 2–3 mm in a 5–10 mm node and false positives may occur if there is fibrosis or fatty replacement of the nodes [75]. It may be combined with other imaging methods such as PET to further increase its sensitivity and specificity.

Developments in MR techniques

One of the well recognized advantages of MRI is its greater functionality in characterizing tissues. With this ability, MR may provide better knowledge of tumour extent through pathophysiology or tumour response compared with simple morphological assessments. This additional information on active tumour regions may be exploited in radiotherapy for the determination of boost volumes, dose escalation, combined therapy with chemotherapy or radiosensitizers, or to select prospective non-responders during a course of radiotherapy for more aggressive treatment.

Cancer growth is usually accompanied by vascular extension and growth in order to meet the nutritional demands of rapid tumour expansion. These tumour features of vascular angiogenesis and increased cellular



(a) (b)

Figure 8. A case example of the use of ultrasmall paramagnetic iron oxide (USPIO) in MRI to evaluate pelvic lymph nodes in a man with prostate cancer. In this case, the lymph nodes (thick arrow) pre-USPIO (a) returned a negative MR signal following USPIO administration (thin arrow (b)) indicating normal lymph node architecture. This was later confirmed on lymph node sampling.

growth are exploited by MR techniques such as MR diffusion and dynamic contrast enhanced MRI. These MR techniques have been available in the research domain for some time and are therefore technically not new developments. However, they are included here as they are not part of standard practice and their role in clinical management is still being assessed.

Dynamic contrast-enhanced MRI

Dynamic contrast enhanced MRI (dcMRI) is a technique which exploits the vascular dynamics inherent in new blood vessel proliferation (neo-angiogenesis) and abnormal vasculature associated with tumour growth. Using fast MR sequences timed to capture the sequential changes in vascular perfusion following injection of low-molecular weight contrast agents, the perfusion or enhancement of the examined tissues and/or organ can be characterized. Time dependent enhancement curves can be subsequently produced. These (qualitative) enhancement curves from different nominated regions of interest (ROI) within the tissues can then be compared and assessed. Different tissue enhancement curves can be recognized for normal, tumour and irradiated tissues [4]. In this manner, ROI exhibiting appropriate enhancement tumour curves can be selected or delineated for radiotherapy targeting or tumour boosting.

The phases of the low-molecular weight contrast agents as they pass through the vascular system and tissues can be assessed by different MR sequences. T_2 weighted sequences are more sensitive to the vascular phase and thus better display tissue perfusion and blood volume effects [76]. T_1 weighted sequences are better at detecting the contrast agents in the extravascular to extracellular space and thus demonstrate microvessel perfusion, permeability and extracellular leakage effects [76]. They can be selected to provide different clinical assessments such as differentiating between benign or normal tissues, localizing active tumour regions and predicting and monitoring tumour response. T_2 methods have been reported to be better suited for evaluating brain tumours, particularly gliomas [77], whereas T_1 methods have been used more widely in breast, musculoskeletal, gynaecological and urological cancers [78–81].

These dcMRI methods have been used to detect tumour recurrence in previously irradiated breast [78] and prostate cancer sites [79, 82] as well as to predict response to radiotherapy in head and neck [83], rectal [80] and cervical cancers [81]. These methods can also be used to assess the nature of lymph nodes where pathologically involved nodes may possess appropriate tumour enhancement patterns on their signal intensity time curves. One limitation is that this technique can only assess a specific region and not all nodal sites.

Diffusion weighted MRI

Diffusion weighted MRI (dwMRI) attempts to assess the diffusion capacity of tissue. This methodology relies on the tumour regions having increased cellular density due to tumour proliferation and therefore results in a reduction of diffusion of water molecules through this abnormal region. Apparent diffusion coefficient (ADC) maps can be generated from different spatial regions of interest. A lower ADC is more likely to contain tumour than a high ADC. Early studies suggest that dwMRI may help distinguish between malignant and benign lesions in the brain [5], but dwMRI studies in other cancer types such as soft tissue sarcomas were not found to be as helpful due to substantial overlap in ADC values [84].

The use of dwMRI tumour defined areas may permit the assessment of tumour response during a course of therapy and regions of poor response may be selected for radiotherapy boosting. Preliminary dwMRI studies in rectal cancer treated using chemoradiation suggest that ADC values may provide indicators of tumour response [85, 86]. This can then be used to optimize treatment strategies during therapy or to initiate adjuvant therapy for the individual patient.

Diffusion tensor imaging

Diffusion tensor MRI (DTI) is a technique that can demonstrate white matter abnormalities based on cerebral tissue anisotrophy (a measure of tissue disorganization) and provide information on brain tumour involvement on white matter tracts. Investigators using DTI suggest that this method may be useful for assessing white matter infiltration by occult tumour [87]. A recent study reported that DTI recorded larger white matter tract abnormalities than seen on T_2 weighted images in 10 out of 13 high-grade gliomas and previously unrecognized contralateral involvement in 4 of these cases [87]. A further planning study suggested that treatment volumes may be optimized using DTI by reducing the PTV compared with CT planning alone and thus provide the opportunity of dose escalation whilst maintaining tissue tolerances [88]. If these findings are confirmed then this information will aid RTP and also provide prognostic information if the extent of invasion is a determinant of disease outcome.

Magnetoencephalography and DTI can also be used in RTP to limit doses to relevant functional regions of the cerebrum or white tracts to reduce specific radiation-induced neurological dysfunction for each patient case to permit plan individualization [89]. DTI methods may offer complementary information to other imaging techniques including PET that are being used for target volume delineation in brain RTP.

Summary

The superior characterization of soft tissues and visualization of tumour extent from MRI can benefit RTP by improving target volume delineation and assessment of planning margins in many cancer subtypes in sites such as the brain, spinal cord, soft tissues of the head and neck, trunk and limbs. In the past decade there have been many advances in MRI technology that can further aid the definition of volumes for both external beam radiotherapy and brachytherapy.

Open and low field MR simulators can provide easier integration of MR for RTP by its lower cost, facilitating

the use of patient immobilization devices and providing fewer image artefacts and less distortion. USPIO contrast agents can evaluate pathological lymph nodes for treatment. Ultrafast volumetric and cine mode sequences can provide temporal assessment of target volume deformity and positioning for image guided radiotherapy.

3 T scanners can provide higher resolution images for better tissue definition and can also benefit MRS applications. MR techniques using dcMR, dwMR and DTI can further assess target volumes with improved and complementary morphological, functional and biological data that can provide the opportunity to nominate biological target volumes and the potential to gauge treatment response. These techniques may also be combined with PET to further increase diagnostic sensitivity and specificity.

This better estimation of target volumes may permit treatment individualization, organ sparing or functional avoidance, strategies of boosting and dose escalation with or without concurrent chemotherapy. All these new advances can increase the scope of MRI in radiotherapy, but currently their role and utility for RTP remains to be defined. However, just as important is the need for specific training for oncologists to understand how to utilize MR images for RTP. There must be close collaboration between diagnostic radiologists, physicists, radiographers and oncologists in order to effectively harness the benefits of MRI for radiotherapy. Ideally, this should be through a dedicated oncology team approach.

Acknowledgments

We are grateful to Dr M Wada, Austin Health Radiation Oncology Centre, for providing the images demonstrating the base of tongue cancer case.

References

1. Mackenzie R, Dixon AK. Measuring the effects of imaging: an evaluative framework. Clin Radiol 1995;50:513–8.
2. Khoo VS. MRI–"magic radiotherapy imaging" for treatment planning? Br J Radiol 2000;73:229–33.
3. Harisinghani MG, Saini S, Weissleder R, Hahn PF, Yantiss RK, Tempany C, et al. MR lymphangiography using ultrasmall superparamagnetic iron oxide in patients with primary abdominal and pelvic malignancies: radiographic-pathologic correlation. AJR Am J Roentgenol 1999;172:1347–51.
4. Padhani AR, Husband JE. Dynamic contrast-enhanced MRI studies in oncology with an emphasis on quantification, validation and human studies. Clin Radiol 2001;56:607–20.
5. Rees J. Advances in magnetic resonance imaging of brain tumours. Curr Opin Neurol 2003;16:643–50.
6. Rohlfing T, Maurer CR Jr, O'Dell WG, Zhong J. Modeling liver motion and deformation during the respiratory cycle using intensity-based nonrigid registration of gated MR images. Med Phys 2004;31:427–32.
7. Padhani AR, Khoo VS, Suckling J, Husband JE, Leach MO, Dearnaley DP. Evaluating the effect of rectal distension and rectal movement on prostate gland position using cine MRI. Int J Radiat Oncol Biol Phys 1999;44:525–33.
8. Sundar S, Symonds RP. Diagnostic radiology for radiotherapist: the case for structured training in cross-sectional imaging (CT and MRI). Clin Oncol (R Coll Radiol) 2002;14:413–4.
9. Board of the Faculty of Clinical-Oncology: Imaging for oncologists: collaboration between clinical radiologists and clinical oncologists in diagnosis, staging and radiotherapy planning. London, Royal College of Radiologists, 2004.
10. Khoo VS, Dearnaley DP, Finnigan DJ, Padhani A, Tanner SF, Leach MO. Magnetic resonance imaging (MRI): considerations and applications in radiotherapy treatment planning. Radiother Oncol 1997;42:1–15.
11. Thornton AF, Sandler HM, Ten Haken RK, McSchan DL, Fraass BA, La Vigne ML, et al. The clinical utility of magnetic resonance imaging in the 3-dimensional treatment planning of brain tumors. Int J Radiat Oncol Biol Phys 1992;24:767–75.
12. Heester MA, Wijrdeman HK, Strukmans H, Witkamp T, Moerland MA. Brain tumor delineation based on CT and MR imaging. Strahlenther Onkol 1993;169:729–33.
13. Sultanem K, Patrocinio H, Lambert C, Corns R, Leblanc R, Parker W, et al. The use of hypofractionated intensity-modulated irradiation in the treatment of glioblastoma multiforme: preliminary results of a prospective trial. Int J Radiat Oncol Biol Phys 2004;58:247–52.
14. Ten Haken RK, Thornton AF, Sandler HM, La Vigne ML, Quint DJ, Frass BA, et al. A quantitative assessment of the addition of MRI to CT-based, 3-D treatment planning of brain tumors. Radiother Oncol 1992;25:121–33.
15. Khoo VS, Adams EJ, Saran F, Bedford JL, Perks JR, Warrington AP, et al. A comparison of clinical target volumes determined by CT and MRI for the radiotherapy planning of base of skull meningiomas. Int J Radiat Oncol Biol Phys 2000;46:1309–17.
16. Mazzara GP, Velthuizen RP, Pearlman JL, Greenberg HM, Wagner H. Brain tumor target volume determination for radiation treatment planning through automated MRI segmentation. Int J Radiat Oncol Biol Phys 2004;59:300–12.
17. Bondiau PY, Malandain G, Chanalet S, Marcy PY, Habrand JL, Fauchon F, et al. Atlas-based automatic segmentation of MR images: validation study on the brainstem in radiotherapy context. Int J Radiat Oncol Biol Phys 2005;61:289–98.
18. Levivier M, Massager N, Wikler D, Goldman S. Modern multimodal neuroimaging for radiosurgery: the example of PET scan integration. Acta Neurochir Suppl 2004;91:1–7.
19. Manavis J, Sivridis L, Koukourakis MI. Nasopharyngeal carcinoma: the impact of CT-scan and of MRI on staging, radiotherapy treatment planning, and outcome of the disease. Clin Imaging 2005;29:128–33.
20. Emami B, Sethi A, Petruzzelli GJ. Influence of MRI on target volume delineation and IMRT planning in nasopharyngeal carcinoma. Int J Radiat Oncol Biol Phys 2003;57:481–8.
21. Chung NN, Ting LL, Hsu WC, Lui LT, Wang PM. Impact of magnetic resonance imaging versus CT on nasopharyngeal carcinoma: primary tumor target delineation for radiotherapy. Head Neck 2004;26:241–6.
22. Lee FK, Yeung DK, King AD, Leung SF, Ahuja A. Segmentation of nasopharyngeal carcinoma (NPC) lesions in MR images. Int J Radiat Oncol Biol Phys 2005;61:608–20.
23. Huch Boni RA, Boner JA, Debatin JF, Trinkler F, Knonagel H, Von Hochstetter A, et al. Optimization of prostate carcinoma staging: comparison of imaging and clinical methods. Clin Radiol 1995;50:593–600.
24. Barentsz JO, Engelbrecht MR, Witjes JA, de la Rosette JJ, van der Graaf M. MR imaging of the male pelvis. Eur Radiol 1999;9:1722–36.
25. Heenan SD. Magnetic resonance imaging in prostate cancer. Prostate Cancer Prostatic Dis 2004;7:282–8.
26. Khoo VS, Padhani AR, Tanner SF, Finnigan DJ, Leach MO, Dearnaley DP. Comparison of MRI with CT for the radiotherapy planning of prostate cancer: a feasibility study. Br J Radiol 1999;72:590–7.

27. Wachter S, Wachter-Gerstner N, Bock T, Goldner G, Kovacs G, Fransson A, et al. Interobserver comparison of CT and MRI-based prostate apex definition. Clinical relevance for conformal radiotherapy treatment planning. Strahlenther Onkol 2002;178:263–8.
28. Roach M, Faillace-Akazawa P, Malfatti C, Holland J, Hricak H. Prostate volumes defined by magnetic resonance imaging and computerized tomographic scans for 3-dimensional conformal radiotherapy. Int J Radiat Oncol Biol Phys 1996;35:1011–8.
29. Kagawa K, Lee WR, Schultheiss TE, Hunt MA, Shaer AH, Hanks GE. Initial clinical assessment of CT-MRI image fusion software in localization of the prostate for 3D conformal radiation therapy. Int J Radiat Oncol Biol Phys 1997;38:319–25.
30. Rasch C, Barillot I, Remeijer P, Touw A, van Herk M, Lebesque JV. Definition of the prostate in CT and MRI: a multi-observer study. Int J Radiat Oncol Biol Phys 1999;43:57–66.
31. Sannazzari GL, Ragona R, Ruo Redda MG, Giglioli FR, Isolato G, Guarneri A. CT-MRI image fusion for delineation of volumes in three-dimensional conformal radiation therapy in the treatment of localized prostate cancer. Br J Radiol 2002;75:603–7.
32. Joon DL, Nguyen B, Khoo V, Joon ML, See A, Wada M, et al. Assessing the impact of MRI on determination of planning volumes for conformal prostate radiotherapy. Int J Radiat Oncol Biol Phys 2005;63(S1):S331.
33. Jackson AS, Parker CC, Norman AR, Padhani AR, Huddart RA, Horwich A, et al. Tumour staging using magnetic resonance imaging in clinically localised prostate cancer: relationship to biochemical outcome after neo-adjuvant androgen deprivation and radical radiotherapy. Clin Oncol (R Coll Radiol) 2005;17:167–71.
34. Lau HY, Kagawa K, Lee WR, Hunt MA, Shaer AH, Hanks GE. Short communication: CT-MRI image fusion for 3D conformal prostate radiotherapy: use in patients with altered pelvic anatomy. Br J Radiol 1996;69:1165–70.
35. Steenbakkers RJ, Deurloo KE, Nowak PJ, Lebesque JV, van Herk M, Rasch CR. Reduction of dose delivered to the rectum and bulb of the penis using MRI delineation for radiotherapy of the prostate. Int J Radiat Oncol Biol Phys 2003;57:1269–79.
36. Menard C, Susil RC, Choyke P, Gustafson GS, Kammerer W, Ning H, et al. MRI-guided HDR prostate brachytherapy in standard 1.5T scanner. Int J Radiat Oncol Biol Phys 2004;59:1414–23.
37. Citrin D, Ning H, Guion P, Li G, Susil RC, Miller RW, et al. Inverse treatment planning based on MRI for HDR prostate brachytherapy. Int J Radiat Oncol Biol Phys 2005;61:1267–75.
38. Polo A, Cattani F, Vavassori A, Origgi D, Villa G, Marsiglia H, et al. MR and CT image fusion for postimplant analysis in permanent prostate seed implants. Int J Radiat Oncol Biol Phys 2004;60:1572–9.
39. Crook J, McLean M, Yeung I, Williams T, Lockwood G. MRI-CT fusion to assess postbrachytherapy prostate volume and the effects of prolonged edema on dosimetry following transperineal interstitial permanent prostate brachytherapy. Brachytherapy 2004;3:55–60.
40. Krempien RC, Daeuber S, Hensley FW, Wannenmacher M, Harms W. Image fusion of CT and MRI data enables improved target volume definition in 3D-brachytherapy treatment planning. Brachytherapy 2003;2:164–71.
41. Haie-Meder C, Potter R, Van Limbergen E, Briot E, De Brabandere M, Dimopoulos J, et al. Recommendations from Gynaecological (GYN) GEC-ESTRO Working Group (I): concepts and terms in 3D image based 3D treatment planning in cervix cancer brachytherapy with emphasis on MRI assessment of GTV and CTV. Radiother Oncol 2005;74:235–45.
42. Joon D, Butcher M, Marr M, Quong G, Feigen M, Wada M, et al. Evaluation of the use of CT planning vs orthogonal films in the treatment of rectal carcinoma. Proceedings of the 12th International Congress of Radiation Research (ICRR). Brisbane, Australia, 2003:258.
43. Khoo V, Joon D, Tan J, Fitt G, Feigen M, Marr M, et al. The utility of multimodality imaging with MRI to determine treatment volumes for chemoradiation in rectal cancer. Eur J Cancer 2005;3:408.
44. Ferri M, Laghi A, Mingazzini P, Iafrate F, Meli L, Ricci F, et al. Pre-operative assessment of extramural invasion and sphincteral involvement in rectal cancer by magnetic resonance imaging with phased-array coil. Colorectal Dis 2005;7:387–93.
45. Urban M, Rosen HR, Holbling N, Feil W, Hochwarther G, Hruby W, et al. MR imaging for the preoperative planning of sphincter-saving surgery for tumors of the lower third of the rectum: use of intravenous and endorectal contrast materials. Radiology 2000;214:503–8.
46. Blomqvist L, Holm T, Nyren S, Svanstrom R, Ulvskog Y, Iselius L. MR imaging and computed tomography in patients with rectal tumours clinically judged as locally advanced. Clin Radiol 2002;57:211–8.
47. Beets-Tan RG, Lettinga T, Beets GL. Pre-operative imaging of rectal cancer and its impact on surgical performance and treatment outcome. Eur J Surg Oncol 2005;31:681–8.
48. Schad LR, Bluml S, Hawighorst H, Wenz F, Lorenz WJ. Radiosurgical treatment planning of brain metastases based on a fast, three-dimensional MR imaging technique. Magn Reson Imaging 1994;12:811–9.
49. Chen L, Price RA Jr, Wang L, Li J, Qin L, McNeeley S, et al. MRI-based treatment planning for radiotherapy: dosimetric verification for prostate IMRT. Int J Radiat Oncol Biol Phys 2004;60:636–47.
50. Lee YK, Bollet M, Charles-Edwards G, Flower MA, Leach MO, McNair H, et al. Radiotherapy treatment planning of prostate cancer using magnetic resonance imaging alone. Radiother Oncol 2003;66:203–16.
51. Finnigan DJ, Tanner SF, Dearnaley DP, Edser E, Horwich A, Khoo VS, et al. Distortion-corrected magnetic resonance images for pelvic radiotherapy treatment planning, In: Faulkner K, Carey B, Crellin A, et al, editors. Quantitative imaging in oncology. London: British Institute of Radiology, 1997:72–6.
52. Tanner SF, Finnigan DJ, Khoo VS, Mayles P, Dearnaley DP, Leach MO. Radiotherapy planning of the pelvis using distortion corrected MR images: the removal of system distortions. Phys Med Biol 2000;45:2117–32.
53. Doran SJ, Charles-Edwards L, Reinsberg SA, Leach MO. A complete distortion correction for MR images: I. Gradient warp correction. Phys Med Biol 2005;50:1343–61.
54. Reinsberg SA, Doran SJ, Charles-Edwards EM, Leach MO. A complete distortion correction for MR images: II. Rectification of static-field inhomogeneities by similarity-based profile mapping. Phys Med Biol 2005;50:2651–61.
55. Petersch B, Bogner J, Fransson A, Lorang T, Potter R. Effects of geometric distortion in 0.2T MRI on radiotherapy treatment planning of prostate cancer. Radiother Oncol 2004;71:55–64.
56. Koch N, Liu HH, Olsson LE, Jackson EF. Assessment of geometrical accuracy of magnetic resonance images for radiation therapy of lung cancers. J Appl Clin Med Phys 2003;4:352–64.
57. Sosna J, Pedrosa I, Dewolf WC, Mahallati H, Lenkinski RE, Rofsky NM. MR imaging of the prostate at 3 Tesla: comparison of an external phased-array coil to imaging with an endorectal coil at 1.5 Tesla. Acad Radiol 2004;11:857–62.

58. Bloch BN, Rofsky NM, Baroni RH, Marquis RP, Pedrosa I, Lenkinski RE. 3 Tesla magnetic resonance imaging of the prostate with combined pelvic phased-array and endorectal coils; Initial experience(1). Acad Radiol 2004;11:863–7.
59. Pirzkall A, Li X, Oh J, Chang S, Berger MS, Larson DA, et al. 3D MRSI for resected high-grade gliomas before RT: tumor extent according to metabolic activity in relation to MRI. Int J Radiat Oncol Biol Phys 2004;59:126–37.
60. Pickett B, Vigneault E, Kurhanewicz J, Verhey L, Roach M. Static field intensity modulation to treat a dominant intra-prostatic lesion to 90 Gy compared to seven field 3-dimensional radiotherapy. Int J Radiat Oncol Biol Phys 1999;44:921–9.
61. Mack A, Wolff R, Scheib S, Rieker M, Weltz D, Mack G, et al. Analyzing 3-tesla magnetic resonance imaging units for implementation in radiosurgery. J Neurosurg 2005;102158–64.
62. Mizowaki T, Nagata Y, Okajima K, Murata R, Yamamoto M, Kokubo M, et al. Development of an MR simulator: experimental verification of geometric distortion and clinical application. Radiology 1996;199:855–60.
63. Deasy NP, Conry BG, Lewis JL, Ford TF, Russell GA, Basu R, et al. Local staging of prostate cancer with 0.2 T body coil MRI. Clin Radiol 1997;52:933–7.
64. Buyyounouski MK, Horwitz EM, Price RA, Hanlon AL, Uzzo RG, Pollack A. Intensity-modulated radiotherapy with MRI simulation to reduce doses received by erectile tissue during prostate cancer treatment. Int J Radiat Oncol Biol Phys 2004;58:743–9.
65. Krauss DJ, Kestin LL, Raff G, Yan D, Wong J, Gentry R, et al. MRI-based volumetric assessment of cardiac anatomy and dose reduction via active breathing control during irradiation for left-sided breast cancer. Int J Radiat Oncol Biol Phys 2005;61:1243–50.
66. Plathow C, Ley S, Fink C, Puderbach M, Hosch W, Schmahl A, et al. Analysis of intrathoracic tumor mobility during whole breathing cycle by dynamic MRI. Int J Radiat Oncol Biol Phys 2004;59:952–9.
67. Liu HH, Koch N, Starkschall G, Jacobson M, Forster K, Liao Z, et al. Evaluation of internal lung motion for respiratory-gated radiotherapy using MRI: Part II-margin reduction of internal target volume. Int J Radiat Oncol Biol Phys 2004;60:1473–83.
68. Ghilezan MJ, Jaffray DA, Siewerdsen JH, Van Herk M, Shetty A, Sharpe MB, et al. Prostate gland motion assessed with cine-magnetic resonance imaging (cine-MRI). Int J Radiat Oncol Biol Phys 2005;62:406–17.
69. Khoo VS, Bedford JL, Padhani A, Leach M, Husband J, Dearnaley DP. Prostate and rectal deformation assessed using cine magnetic resonance imaging (MRI) during a course of radical prostate radiotherapy. Radiother Oncol 2002;64:285.
70. McBain CA, Sykes JS, Buckley DL, Amer A, Moore CJ, Cowan RA, et al. Optimising bladder radiotherapy: MR assessment of time-dependent organ motion. Clin Oncol (R Coll Radiol) 2003;15:13.
71. Mack MG, Balzer JO, Straub R, Eichler K, Vogl TJ. Superparamagnetic iron oxide-enhanced MR imaging of head and neck lymph nodes. Radiology 2002;222:239–44.
72. Portaluri M, Bambace S, Perez C, Giuliano G, Angone G, Scialpi M, et al. Clinical and anatomical guidelines in pelvic cancer contouring for radiotherapy treatment planning. Cancer Radiother 2004;8:222–9.
73. Martin J, Joon DL, Ng N, Grace M, Van-Gelderen D, Lawlor M, et al. Towards individualised radiotherapy for stage I seminoma. Radiother Oncol 2005;76:251–6.
74. Gregoire V, Coche E, Cosnard G, Hamoir M, Reychler H. Selection and delineation of lymph node target volumes in head and neck conformal radiotherapy. Proposal for standardizing terminology and procedure based on the surgical experience. Radiother Oncol 2000;56:135–50.
75. Kim JY, Harisinghani MG. MR imaging staging of pelvic lymph nodes. Magn Reson Imaging Clin N Am 2004;12:581–6.
76. Neeman M, Provenzale JM, Dewhirst MW. Magnetic resonance imaging applications in the evaluation of tumor angiogenesis. Semin Radiat Oncol 2001;11:70–82.
77. Sugahara T, Korogi Y, Kochi M, Ikushima I, Hirai T, Okuda T, et al. Correlation of MR imaging-determined cerebral blood volume maps with histologic and angiographic determination of vascularity of gliomas. AJR Am J Roentgenol 1998;171:1479–86.
78. Dao TH, Rahmouni A, Campana F, Laurent M, Asselain B, Fourquet A. Tumor recurrence versus fibrosis in the irradiated breast: differentiation with dynamic gadolinium-enhanced MR imaging. Radiology 1993;187:751–5.
79. Hawnaur JM, Zhu XP, Hutchinson CE. Quantitative dynamic contrast enhanced MRI of recurrent pelvic masses in patients treated for cancer. Br J Radiol 1998;71:1136–42.
80. De Vries A, Griebel J, Kremser C, Judmaier W, Gneiting T, Debbage P, et al. Monitoring of tumor microcirculation during fractionated radiation therapy in patients with rectal carcinoma: preliminary results and implications for therapy. Radiology 2000;217:385–91.
81. Mayr NA, Yuh WT, Arnholt JC, Ehrhardt JC, Sorosky JI, Magnotta VA, et al. Pixel analysis of MR perfusion imaging in predicting radiation therapy outcome in cervical cancer. J Magn Reson Imaging 2000;12:1027–33.
82. Rouviere O, Valette O, Grivolat S, Colin-Pangaud C, Bouvier R, Chapelon JY, et al. Recurrent prostate cancer after external beam radiotherapy: value of contrast-enhanced dynamic MRI in localizing intraprostatic tumor--correlation with biopsy findings. Urology 2004;63:922–7.
83. Tomura N, Omachi K, Sakuma I, Takahashi S, Izumi J, Watanabe O, et al. Dynamic contrast-enhanced magnetic resonance imaging in radiotherapeutic efficacy in the head and neck tumors. Am J Otolaryngol 2005;26:163–7.
84. Einarsdottir H, Karlsson M, Wejde J, Bauer HC. Diffusion-weighted MRI of soft tissue tumours. Eur Radiol 2004;14:959–63.
85. Dzik-Jurasz A, Domenig C, George M, Wolber J, Padhani A, Brown G, et al. Diffusion MRI for prediction of response of rectal cancer to chemoradiation. Lancet 2002;360:307–8.
86. Kremser C, Judmaier W, Hein P, Griebel J, Lukas P, de Vries A. Preliminary results on the influence of chemoradiation on apparent diffusion coefficients of primary rectal carcinoma measured by magnetic resonance imaging. Strahlenther Onkol 2003;179:641–9.
87. Price SJ, Burnet NG, Donovan T, Green HA, Pena A, Antoun NM, et al. Diffusion tensor imaging of brain tumours at 3T: a potential tool for assessing white matter tract invasion? Clin Radiol 2003;58:455–62.
88. Jena R, Price SJ, Baker C, Jefferies SJ, Pickard JD, Gillard JH, et al. Diffusion tensor imaging: possible implications for radiotherapy treatment planning of patients with high-grade glioma. Clin Oncol (R Coll Radiol) 2005;17:581–90.
89. Aoyama H, Kamada K, Shirato H, Takeuchi F, Kuriki S, Iwasaki Y, et al. Integration of functional brain information into stereotactic irradiation treatment planning using magnetoencephalography and magnetic resonance axonography. Int J Radiat Oncol Biol Phys 2004;58:1177–83.

Clinical Oncology 22 (2010) 688–697

Contents lists available at ScienceDirect

Clinical Oncology

journal homepage: www.elsevier.com/locate/clon

Overview

Computed Tomography–Magnetic Resonance Image Registration in Radiotherapy Treatment Planning

J.N.H. Brunt

Physics Department, Clatterbridge Centre for Oncology NHS Trust, Wirral, Merseyside, UK

Received 26 April 2010; accepted 28 June 2010

Abstract

Magnetic resonance imaging (MRI) is being increasingly used in radiotherapy treatment planning (RTP). MRI has the potential to provide improved localisation of target volumes, leading to better tumour control rates and reduced normal tissue complications, due to capabilities including excellent soft-tissue discrimination and the ability to provide scans in which the image contrast is weighted according to different tissue properties. When computed tomography (CT)–MRI image registration is deployed, MR's advantages are combined with CT's geometrical security and its ability to provide electron density information. The quality of CT–MRI image registration can be favourably influenced by aspects of scan acquisition, including patient positioning/immobilisation and scan protocols. Appropriate protocols can ameliorate the possible presence of MR spatial distortions and other artefacts, but quality assurance of scanning remains essential. Here, the methods and quality assurance of CT–MR image registration are discussed. Developments in MRI scanner technology are progressively offering advantages for RTP, in terms of the possibility of better matching of patient positioning versus CT in a greater range of anatomical regions, while allowing thinner slices for better image quality in reformatted orthogonal planes.

Key words: Image fusion; image registration; MRI; radiotherapy planning

Statement of Search Strategies Used and Sources of Information

The following databases were searched between 2009 and April 2010: Medline, later Scopus, which includes all Medline titles (no restrictions set on date range, document type or subject areas) using the following search terms: 'registration' and 'MR*' and 'radiotherapy'; 'registration' and 'MR*' and 'radiation'; 'image' and 'registration' and 'quality assurance'; 'fusion' and 'MR*' and 'radiotherapy'; 'fusion' and 'MR*' and 'radiation'; 'magnetic resonance' and 'treatment planning'; 'image' and 'registration' and 'radiation'.

Introduction

Magnetic resonance imaging (MRI) and computed tomography (CT) are being increasingly used in radiotherapy treatment planning (RTP). MRI may inform the RTP process simply by having MR images displayed alongside RTP computer displays, for example on a Picture Archiving and Communication System (PACS) monitor. MRI is used formally, however, in RTP by transferring image information on to RTP computers. A review by Evans [1] provides an overview of the roles in radiotherapy of CT and MRI, among other anatomical imaging modalities. An International Atomic Energy Agency report by Huq *et al.* [2] classified conformal therapy according to the methodology and tools deployed and indicated that for level 3 advanced three-dimensional conformal radiotherapy (the highest level of accuracy using intensity-modulated radiotherapy or stereotactic radiosurgery), the associated imaging system would be co-registered CT with MR or positron emission tomography. MR has a number of properties that should be considered with regard to its use in RTP. It has the potential to provide improved localisation of target volumes, leading to better tumour control rates and reduced normal tissue complications, because of its: (a) superior soft-tissue discrimination capabilities even in comparison with CT (see Fig. 1), (b) ability to provide

Author for correspondence: J.N.H. Brunt, Medical Physics Department, Clatterbridge Centre for Oncology, NHS Foundation Trust, Bebington, Wirral, Merseyside CH63 4JY, UK. Tel: +44-151-334-1155x4777; Fax: +44 151-482-7860.

E-mail address: John.Brunt@ccotrust.nhs.uk

0936-6555/$36.00
doi:10.1016/j.clon.2010.06.016

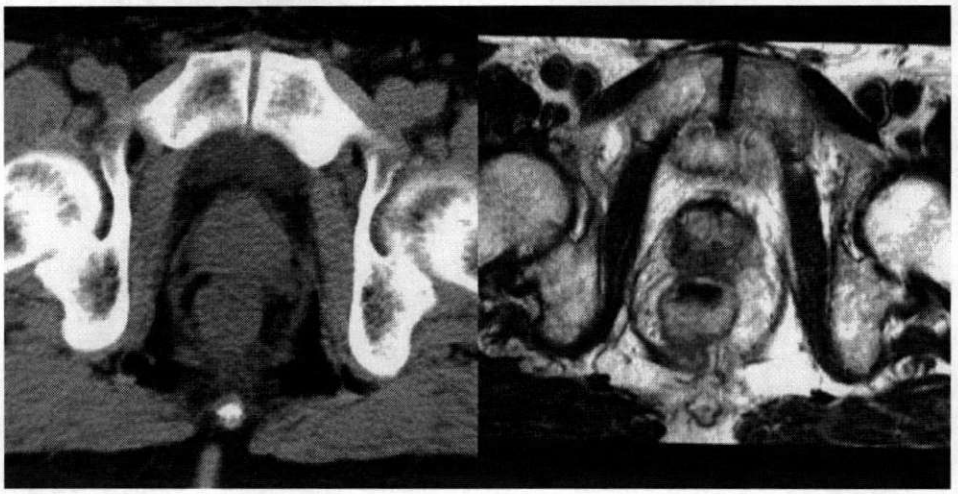


Fig. 1. Computed tomography (left) and magnetic resonance (right) pelvic images used for planning prostate radiotherapy. The superior soft-tissue contrast of magnetic resonance is evident.

scans in which the image contrast is weighted according to different MR tissue properties, including spin–lattice relaxation time (T1), spin–spin relaxation time (T2) and proton density, (c) ability to image in any plane, allowing evaluation of superior and inferior margins, and (d) lack of bony artefacts. MR's inability to provide electron density information is not a problem when MR can be used in combination with CT, although complications for combination are that in MR solid bone is visualised only by the absence of signal (but marrow is visualised), and that various spatial distortions and artefacts may be present.

RTP based only on MR, without a planning CT, has been discussed [3,4], and because MRI involves no ionising radiation, this would eliminate the ionising radiation dose that CT delivers (including dose to areas that will not receive dose during treatment). However, in view of the above-mentioned limitations of MRI and with access to CT scanning being generally better than access to MR, the usual practice, when MR is available for planning scans, remains to use MR scans in conjunction with CT planning scans.

Here, scan acquisition issues affecting subsequent CT–MR image combination quality will be described — these include patient positioning/immobilisation and scan protocols. Because scanners, PACS and RTP work-stations all implement DICOM (Digital Imaging and Communications in Medicine) standards, the transfer of image and header data is now usually straightforward. The combination of MR with CT data can be considered to consist of two aspects, namely aligning the modalities spatially (image 'registration') and merging information from the aligned voxels (image 'fusion'). This terminology is normally used in the formal literature, but system manufacturers often use the term fusion to describe software that has registration as its primary purpose. Displaying merged or blended information from the aligned voxels is, however, a useful aid to verification of registration.

Scan Acquisition Issues Affecting Subsequent Registration Quality

The feasibility and difficulty of image registration is heavily influenced by the similarity, for the CT and MR scans, of patient positioning achieved, and therefore related issues are considered here.

GENERAL ISSUES CONCERNING PATIENT POSITIONING AND IMMOBILISATION

Planning CT scans are normally obtained using a flat patient support. The standard patient support for some MR scanners is also flat, but for many it is concave. In the latter case, it should be feasible for a hospital workshop to manufacture a flat table top insert for the MR patient support.

MR-compatible knee rests and ankle stocks can rest on the flat table top and, where appropriate, can help to replicate positioning for CT and MR scans.

PATIENT POSITIONING FOR BRAIN SCANS

Standard head coils are frequently used for RTP MR brain scans. CT–MR image registration will be improved if neck flexion is similar in both modalities. Therefore, for the MR scan, MR radiographers seek to replicate the head angulation for CT (this may be visualised in a lateral digital photograph of the patient in the immobilisation device). The MR scan is then carried out with zero angulation of the MR slice stack about the patient's right–left axis. Skull-fixed stereotactic frames may impose a more 'chin-down' head angulation and, therefore, the MR scans may advantageously be carried out with the head coil on a wedge, and additionally several degrees of electronic slice angulation about the right–left axis may be applied. In most cases, this approach produces MR slice sets whose angulation approximates that of the (non-tilted) CT scans obtained

with the frame in place. Inter-scan angulation differences exceeding 20° are rarely encountered with this approach.

Feasibility of using Radiotherapy Treatment Planning Head Immobilisation Devices for Magnetic Resonance Brain Scans

The following paragraphs discuss why it is common not to use RTP head immobilisation devices for MR brain RTP scans.

Treatment shells/masks

Standard MR head coils are normally too small to accommodate plastic treatment shells. However, see below for comments concerning the use of phased array surface coils for extracranial RTP scans.

Relocatable stereotactic frames

When better patient immobilisation than is achievable with plastic treatment shells/masks is required, a relocatable metallic frame may be used for CT, but the RTP MR scan is carried out in the MR head coil, without the frame.

Skull-fixed non-relocatable stereotactic frames

When treating in these frames, sharp-tipped screws fix the ring securely to the skull, and the CT scan is obtained with the frame in place. The MR scan is obtained before the treatment day, to reduce the time from fixation to treatment. Hence, the issue of possibly using the frame in MR does not arise. No lateral photograph of the head in the frame can be available at the time of the MR scan.

Components of the frame typically propagate streak artefacts into some CT transaxial slices, and this is a further motivation for including MR, particularly if parts of the lesion are obscured by artefact in CT (for some locations of pathology, for example adjacent to the skull base, in CT there will probably be streak artefact propagated from dense bone, also).

OPTIMISING SIMILARITY OF TISSUE POSITIONING IN COMPUTED TOMOGRAPHY AND MAGNETIC RESONANCE SCANS OF VARIOUS ANATOMICAL AREAS

Extracranial head and neck

When the pathology to be treated extends below the base of the skull, obtaining similarity of neck flexion becomes a more critical issue for accurate registration.

Immobilisation shells extending to the shoulders may be used during CT scanning when pathology extends into the neck. Such shells can be made compatible with MR, provided that non-metallic base frames, shell-attachment plates, nuts, bolts and other fittings are used. It is advantageous if the available complete immobilisation shell and base will fit in the scanner bore [5]. The base frames of some commonly used shells, however, have a right–left dimension exceeding 60 cm, and are therefore too large for the vast majority of current MR scanners (which are of the cylindrical type, and have a bore diameter not exceeding 60 cm at the magnet isocentre). The immobilisation shell may incorporate a patient-specific, vacuum-formed neck and shoulder rest (a 'vac bag').

A helpful compromise is to deploy in MR the plastic head rest from the immobilisation device, together with a flat table top (and MR-compatible plastic knee support for patient comfort). If the 'vac bag' is not too wide to fit into the MR scanner bore, and is MR-compatible at the field strength used, it can usefully also be used in conjunction with the flat table top (see Fig. 2).

The MR coils used would typically be phased array coils, in various configurations such as the following. A pair of circular surface coils (typical radius 20 cm), usually deployed anteriorly and posteriorly (beneath the flat table top). Left and right coil positioning can alternatively be useful, but the coils have to be placed more superiorly, to avoid the shoulders. Alternatively, surface body coils can be deployed posteriorly (again beneath the flat table top) and anteriorly.

There is also a trend towards scanners whose cylindrical bore diameters exceed 60 cm (see later section), which will allow the aforementioned shoulder-accommodating base frames to enter the scanner.

Thoracic and lumbar spine

As with the neck, obtaining similarity of flexion between vertebrae is a key issue. It is helpful for the MR scan to be carried out using a flat table top, for knee and ankle rests to be used if they are used for CT, and for a 'vac bag' cradle to be used. For some treatments, including those for pathology of the sacrum and coccyx, the planning CT would usually be carried out prone, and this positioning should be repeated for MR. MR surface array coils would typically be deployed.

Pelvis

MR-compatible knee rests and ankle stocks are used on a MR flat patient support, with the posterior elements of a set of phased array coils beneath it. MR positioning is based on placing the anatomical location of the pathology at the magnet centre, rather than on skin markers, so wall mounted lasers in MR are not essential. Both CT and MR scans are obtained with extensive coverage in the inferior–superior direction (up to ~170 mm). Patient preparation may involve an attempt to achieve similarity of bladder and bowel filling between scans, as dissimilarity can adversely affect image registration [6].

ISSUES CONCERNING SCAN PROTOCOLS

Details of scan protocols significantly affect registration quality and are, therefore, briefly discussed here.

A water fat (spatial) shift (WFS) results from the different resonant frequencies of materials, affecting tissues according to their composition. This scan parameter should not exceed a maximum value (e.g. 1.0 mm) in each anatomical region. With reduced WFS (increased bandwidth), noise is collected over a greater range of radiofrequencies, reducing image quality (signal to noise ratio). To compensate, longer duration scans are often required. Radiofrequency heat deposition should not be permitted to exceed recommended time-averaged specific absorption rate (SAR) values for safety [7], even with longer duration sequences.

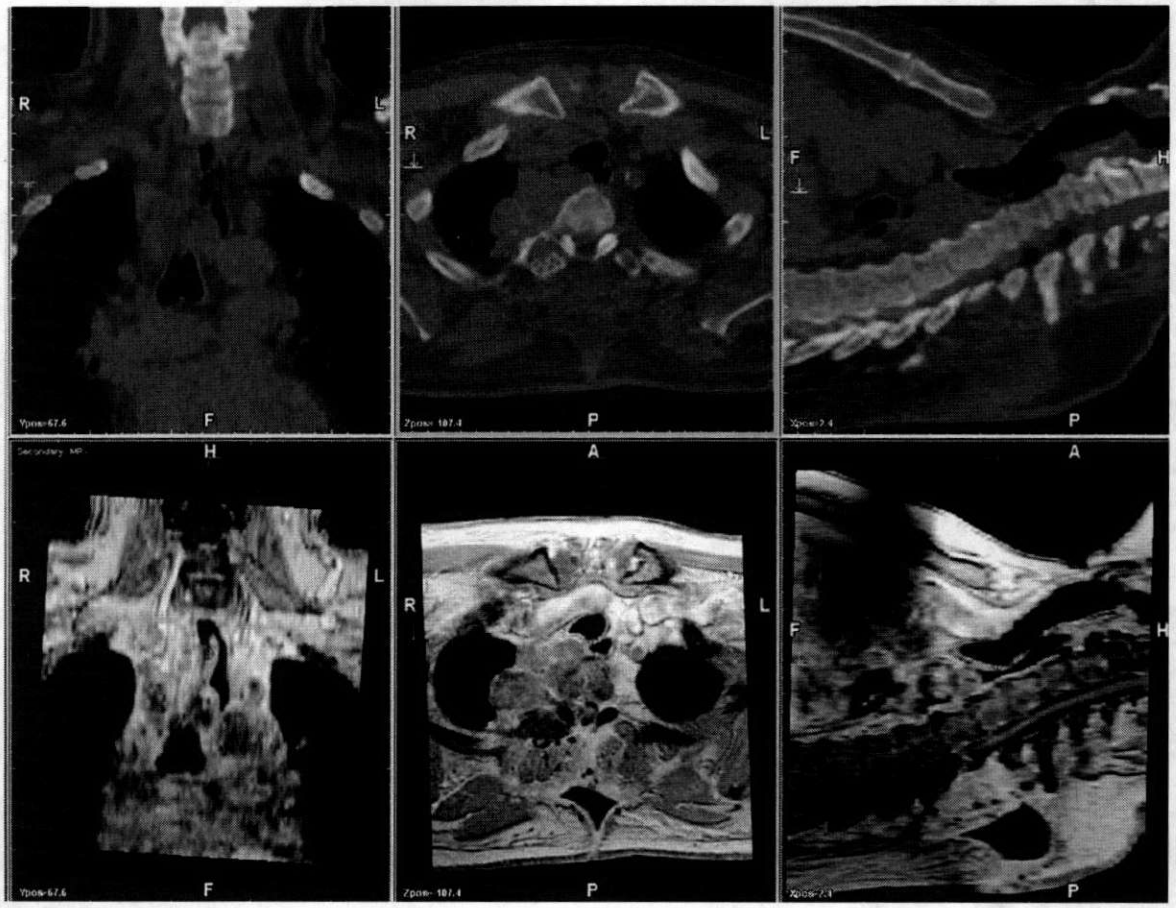


Fig. 2. Registered computed tomography (top row) and magnetic resonance (bottom row) scans. Coronal and sagittal reformats are left and right of the original transaxial scans. In magnetic resonance, a flat table top, a plastic head rest and a 'vac bag' shoulder/neck rest were used.

Table 1 provides example MR parameters for various planning applications. An infinity of possible combinations of MR parameter values exists, with no single perfect combination. The presented values relate to the commonly used 1.5 T field strength, but vary between centres, according to field strength, scanner manufacturer/model and type of receiver coil. Also, centres have different preferences regarding in which anatomical areas to use different scan techniques (spin echo, gradient echo, etc). Unlike Hounsfield values from CT, the numerical pixel values produced by most MR sequences (e.g. a T1-weighted sequence) will vary not only between scanner makes and models, but also according the exact scanning parameters. For a specific type of sequence, however, there will be substantial similarity of tissue contrast from different scanners. Compared with diagnostic scans, RTP scans may use thinner slices, which may be compensated by some reduction in matrix size.

Quality Assurance of Magnetic Resonance Scans

MR scanners should undergo regular quality assurance (QA) testing [8]. One QA test type of particular relevance for RTP is spatial distortion. It was reported [6] that the spatial distortion of a Philips 1.5 T scanner, measured using conventional QA phantoms for RTP planning sequences, increases with distance from the magnet's isocentre, but is typically less than 1 mm and less than 2 mm within radii of 10 and 15 cm, respectively. These figures are satisfactory, respectively, for RTP of brain and of extracranial head and neck or of pelvic structures visualised in small fields of view (FOVs) (see Fig. 3). Broadly similar values on a Siemens 1.5 T scanner have been reported [5].

It is reasonable to expect that on most MR scanners spatial distortion will, however, be unacceptably large for RTP purposes at the pelvic skin surface, which is considerably further than these distances from the magnet isocentre. Within the small FOVs used for pelvic RTP these large distortions may not be encountered.

The trend (see below) to shorter MR magnets (to improve patient comfort) presents challenges to scanner manufacturers to maintain acceptably low spatial distortion. This makes QA particularly important for these magnets.

Quality assurance should also include an examination of slice width and profile. Where three-dimensional (thin-slice) datasets are being considered, this is especially

Table 1
Typical magnetic resonance parameters for some sequences commonly used for radiotherapy treatment planning magnetic resonance scans in various anatomical regions. All sequences in this table are for the transaxial plane. The final lines of the table contain some computed tomography parameter information for comparison

Anatomical region	Brain (non-stereotactic)		Brain (stereotactic)		Head and neck†		Pelvis‡
Receiver coil type*	Quadrature head		Quadrature head		Pair of phased array surface coils		Phased array body surface coils (four elements in total)
	T2 W TSE	T1 W multi-slice GE§	T2 W TSE	T1 W 3D GE§	T2 W TSE	T1 W multi-sliceGE§	T2 W TSE
Magnetic resonance parameter							
FOV	230	230	230	230	230	230	220
Rectangular FOV (%)	80	80	80	80	80	80	100
Matrix of scan (of reconstruction)	384 (512)	256	256	256	384 (512)	256	256
Number of slices	60	60	90	150	75	75	56
Slice thickness/gap (mm)	3.0/0.0	3.0/0.0	2.0/0.0	2.0/−1.0	3.6/0.0	3.6/0.0	4.0/−1.0
TR (ms)	6100	500	3075	25	7750	600	2900
TE (ms)	100	4.3	100	4.6	100	3.9	90
Flip angle if not 90°		80		30		80	
Echo train length (where applicable)	22		22		22		14
Scan duration (min)	4.7	5.6	6.2	6.6	5.9	6.5	5.5
Water fat shift (pixels)	1.0	1.0	0.7	1.0	1.0	1.0	1.15
Computed tomography							
Slice thickness/gap (mm)‖	3.0/0.0		2.0/−1.0		3.0/0.0		3.0/0.0

SE, spin echo; TSE, turbo(fast) spin echo; GE, gradient echo; TR, repetition time; TE, echo time; T1 W, T1 weighted; T2 W, T2 weighted; FOV, field of view.

* Most centres will have receiver coils with at least as many elements as the ones mentioned; where coils with more elements in a phased array are available, alternative compromises of scan parameter values may be appropriate, for example to reduce scan durations.

† Slice separation is here shown as exceeding 3 mm, as this may be required to achieve sufficient superior–inferior coverage within an acceptable time duration of scanning. Where shorter coverage is sufficient, the slice gap may be made negative, thus providing smaller slice separation, with over-contiguous slices.

‡ This uses a relatively small FOV, not normally extending to lateral skin surfaces.

§ Repeated pre- and post-contrast medium.

‖ The number of computed tomography slices is not explicitly stated here and clearly will depend on the coverage required along the superior–inferior axis, but typically 60 or more slices would be obtained, with 180 or more for stereotactic cases.

important, as an informed decision should be made concerning the use of sequences whose slice profiles have sidelobes (sinc function like), which could lead to structures (normal or pathological) appearing (with reduced contrast) in slices away from their true locations.

Carrying out RTP based only on MR, without a planning CT, has been mentioned above, but raises some issues.

The planning CT can normally be regarded as geometrically secure, and therefore provides a template against which any geometrical distortion in the MR may be revealed. A comparison of CT and MR images, once registered as described below, therefore provides an additional opportunity for QA of the geometrical reliability of MR scans. In particular, poor inter-modality matching of anatomy, either overall or locally, can provide evidence of MR distortion. The probable magnitude of the MR system distortion close to the magnet's isocentre has been mentioned above, but additional MR distortion is produced by the patient. In particular, different tissues vary in their magnetic susceptibility and this leads to distortion of the magnetic field within the patient and, hence, image distortion. This will be particularly apparent at for tissues adjacent to boundaries between different materials, for example bone and air.

Some patients may be excluded from MR scanning due to the presence within them of ferromagnetic or non-ferromagnetic metals, but others may be scanned despite the presence of metals (for example, tooth fillings). Typically, there will be localised geometrical distortion at these locations.

Correction of both system- and patient-induced distortion has been an area of research [9], including in volumes as large as the pelvis [10,11]. Correction may involve substantial QA using sophisticated test objects to characterise distortion. Distortion correction can involve disadvantages such as longer scan times and some loss of image quality. Such correction has not yet been widely incorporated into routine clinical practice.

For the present time, RTP based on MR registered to geometrically secure CT remains the mainstream approach.

Having warned of the inevitable existence of some degree of distortion in MR, it should be re-emphasised that with most MR scanners, provided that the scanner has been properly set up and maintained, system distortion

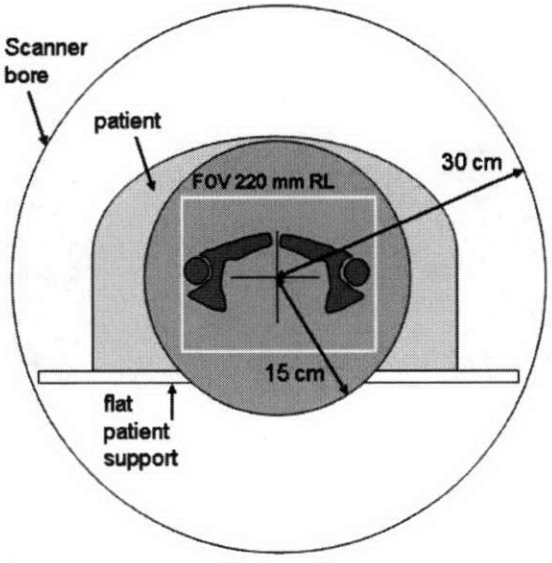


Fig. 3. Schematic diagram of relevant dimensions for pelvic radiotherapy treatment planning. The central box represents the range of small fields of view, 180 mm up to 220 mm, often used as part of a magnetic resonance pelvic examination, in comparison with a 15 cm radius circle, within which magnetic resonance spatial distortion is likely to be less than 2 mm (must be checked by local quality assurance). Distortion at the skin surface would probably be much greater than 2 mm.

close to the magnet centre may be sufficiently limited that it does not present an obstacle to the use of MR for RTP, where the FOVs involved are small (as described above). It is, however, imperative that QA has been carried out to ensure that distortion, with a specific scanner and a particular imaging sequence, is not of a severity incompatible with such use.

Trends in Scanner Design Relevant to Radiotherapy Treatment Planning

For around 15 years, the typical cylindrical MR scanner has had a bore diameter no larger than 60 cm at the centre of the magnet. New models of MR scanners are now following the trend towards larger bore diameters that CT scanners have already followed.

A few years ago, a manufacturer introduced a 1.5 T scanner with a 70 cm minimum bore diameter. This was a favourable development for patients who might experience claustrophobia in smaller-bore scanners, and for patient comfort in general. A specific advantage for RTP was that a greater range of immobilisation devices can fit within the scanner, and patient positioning can be more flexible. Other manufacturers have within approximately the last year introduced 1.5 T scanners with 70 cm bores. Two manufacturers now also offer 3 T, 70 cm bore scanners. Buyers should always ensure that they have unequivocally determined the true minimum bore diameter of any scanner considered.

There is also a trend towards shorter bore lengths in MR scanners, from the 1.8 m that has been typical recently, to less than 1.5 m in the latest models. This again improves patient comfort, as the patient's head may be closer to, or outside, the end of the bore for examination of some areas of the body. In contrast, the trend in multi-slice CT scanners towards ever larger numbers of simultaneously obtainable slices, tends to prevent shorter CT scanner designs being developed. A *de facto* new standard for MR scanner bore dimensions appears to be settling in the region of: length just less than 1.5 m, minimum bore diameter about 70 cm. For scanners with bore lengths significantly less than 1.5 m, achieving acceptably low spatial distortion over a substantial range of coverage in the cranio-caudal direction may be challenging. An interesting approach for obtaining optimal image quality is to utilise a continuously moving patient support (similar in this respect to CT), and always to collect MR data at the magnet centre, however long the overall cranio-caudal coverage.

The high levels of acoustic noise generated by the gradient coils of MR scanners often require patients to use ear plugs or defenders, and are an obstacle to communication between radiographers and patients during scans. In some scanners, acoustic noise has been massively reduced by vacuum encasement of the gradient coils, thus eliminating these problems and allowing patients to be more relaxed and potentially better able to avoid unwanted movement during scans.

The trend in MR surface coils for signal collection is towards increasing numbers of (individually smaller) coil elements, so the set of coils may have as few as two elements or could have dozens of elements. Some MR scanner manufacturers are introducing arrays of coils that are normally left in place beneath the patient support surface, and appropriate combinations of coil elements can then be used with elements of anterior body coils, to optimise signal collection. This optimisation allows thinner image slices for better image quality in images reformatted into orthogonal planes for RTP purposes.

New MR scanning sequences involving motion artefact reduction techniques have the potential to allow clear images to be acquired even when there is patient motion; this may be of particular usefulness for longer duration planning scans.

Scanners of 'open' configuration have lower field strength than the 1.5 T that is common for cylindrical scanners. The lower field strength leads to a reduction in image quality and incapability to carry out some of the more advanced types of scan. One manufacturer does, however, offer a 1.0 T open scanner (using a superconducting magnet) and, where it can be afforded, this provides image quality comparable with the commonly used 1.5 T cylindrical models, in combination with exceptional flexibility of patient positioning.

For MR scanners used for RTP, the most common field strength remains 1.5 T. It has become normal for neurology/neurosurgery centres to have 3 T scanners, and this field strength is coming into increasing use for more general MR scanning. Stanescu *et al.* [4] measured spatial distortion of about 4 mm for a spherical volume of radius 10 cm in a 3 T

scanner, necessitating a distortion correction process as a prerequisite for RTP. The improved image quality at 3 T is appealing for application to RTP. By forgoing some of the image quality improvement, scans can be of shorter duration than at 1.5 T, offering the possibility of reducing artefacts from physiological or other forms of motion. The advantages of 3 T imaging must be weighed against other drawbacks including the need to work around increases in SAR values, WFS and susceptibility artefact. The considerable cost premium for 3 T scanners remains a substantial obstacle to their more general deployment.

Research and development has been reported during the last decade concerning different approaches to integration of a radiotherapy treatment device and an MR scanner into a single device that would visualise pathology at the time of treatment [12–15]. This would expedite dynamic adjustment of a treatment plan during a course of therapy.

Registration of Computed Tomography and Magnetic Resonance Scans

Even if there is acceptably little distortion in an MR scan, it can be expected to differ geometrically from a CT scan of the same anatomical area, in slice angulation and coordinate origin position, as well as the number, thickness and spacing of slices, and the image matrix size and pixel dimensions, hence necessitating image registration.

There have been reviews of the literature on medical image registration by Maintz and Viergever [16] and Hill *et al.* [17], which include further references to several hundred papers in the field. These concern research into registration methods of different categories. Specifically of current interest is the application to RTP, and here the vast majority of registrations conducted fall into the category of 'rigid body' transformations, involving shifting the set of MR slices in x, y, z and rotating it about the x, y, z axes (pitch, yaw, roll) to match the CT. Usually the system needs to interpolate data between slices of the MR, to display an MR slice matching the position and angulation of each CT slice.

Automated Methods

Commercial systems provide automatic registration, in which a computer program brings the modalities into alignment. The most successful current automated algorithms [18] seem to use 'mutual information' techniques, examining voxel density distributions,. The increasing speed of computers now allows such techniques to produce results in considerably less than 1 min.

Modern registration software typically allows the user to constrain the volume within which the automated method will operate, for example by defining a 'bounding box'. In this way, anatomical regions where there could be significant, but irrelevant, discrepancies between CT and MR, can be excluded. For example, the external ear in a patient being planned for brain radiotherapy.

Quality assurance of automated image registration is discussed in a later section.

Manual Methods

Commercial systems also provide manual registration methods, in particular points-based and interactive methods.

A classical method of registration is to use fiducial points (e.g. anatomical landmarks). Whereas three points identified on each modality are the minimum to define the rigid body transformation, at least eight are commonly used, so that the registration will be less at risk of error due to inexact placement of an individual point. A key reference to this method is Hill *et al.* [19], who report the use of anatomical landmark registration in 35 patients and provide lists of anatomical landmarks.

Particularly useful locations of brain landmarks replicated to both left and right of image datasets are the vestibulocochlear nerve and other spatially restricted structures of the inner ear, the zygomatic bones and, occipitally, the lambdoid suture, while mid-sagittal locations include the basilar artery, superior lobes of the cerebellum, and the junction of the straight sinus and the great vein of Galen. In addition to these, the centres of the eye globes are often used among the points when getting an initial approximate registration. Although skin markers containing CT and MR visible contrast can provide alternative starting points, movement, e.g. stretching, of skin with respect to underlying tissue can make reproducibility of their positions unreliable.

Interactive manual registration typically involves a colourwash-tinted MR dataset and a greyscale (or complementary colourwash) CT dataset. These are displayed superimposed while the user applies shifts and rotations to the MR by manipulation of the computer mouse, or of on-screen sliders for the individual movements. The registration should then be checked using anatomical landmarks.

Image Registration: Future Developments

Non-rigid or deformable registration, although continuing to be a field of active research and development, remains relatively little used in clinical RTP. This is largely due to the difficulty of achieving adequate validation of its reliability, although some progress has been achieved [20].

A promising approach in anatomical areas where a deformable registration may be highly advantageous, would be to carry out rigid body registrations of multiple regions of interest (mRoI). In the head and neck, these would cover specific vertebrae and other bony structures, and image deformation would be applied to bring the specific registered mRoI into an overall registration. An mRoI method has been investigated [21] in head and neck CT–CT registration.

Quality Assurance of Image Registration

It is helpful to have numerical values for the consistency of the registration over the volume of interest. This information has been provided for the set of fiducial points (and

for the individual points with respect to the set) by systems supporting this method.

Failure of automated registration may be severe and hence obvious, but can also be subtle. A subtle failure of, say, 2 mm would not necessarily matter for some treatments, but would be completely unacceptable for others. Anatomical landmarks should be used to check automated registrations. Particular care should be exercised if there has been a significant time interval between MR and CT scans, as morphological changes may have occurred, which could render an automated registration inaccurate. For example, the size of ventricles in the brain could change. A solid immobilisation shell might mimic skin surface in CT and cause inaccuracy in automated registration.

Criteria for the selection of landmarks (see Fig. 4) for the verification of automated techniques are essentially the same as for landmark-based manual registration. In each case, to minimise the possibility of the existence of errors in the rotation as well as the translation (x, y, z shifts) between the modalities, it is important to select landmarks: (1) that can be accurately localised in three dimensions (accuracy in the cranio-caudal direction normally being the most difficult to achieve with transaxial datasets), (2) widely spaced around the imaged anatomy, with ones as far anterior, posterior, superior, inferior, right and left as feasible. Localised areas of calcification (within trabecular bone or elsewhere), when they provide signal voids in MR, can sometimes constitute landmarks that fulfil the first criterion particularly well.

A facility that expedites CT–MR landmark checking is the linked cursor (displayed in corresponding, linked locations on both modalities). The cursor should be simultaneously shown on two (and preferably three) views (selectable from transaxial, coronal and sagittal) on both modalities. Alternative facilities in this context are split screen and chequerboard displays [22].

Some software systems provide numerical values for the shifts and rotations comprising the rigid body registration. Where this information is available, numerical comparison of different deployed methods is expedited. For example, an interactive shift and rotate manual method can be deployed, and the parameters recorded. An automated method can then be deployed, starting either from the initial positions of the CT and MR datasets, or from the results of the manual registration, and the parameters again recorded. It may be perceived that the automated registration has provided either an improvement or a degradation of registration quality. Review of which, if any, of the

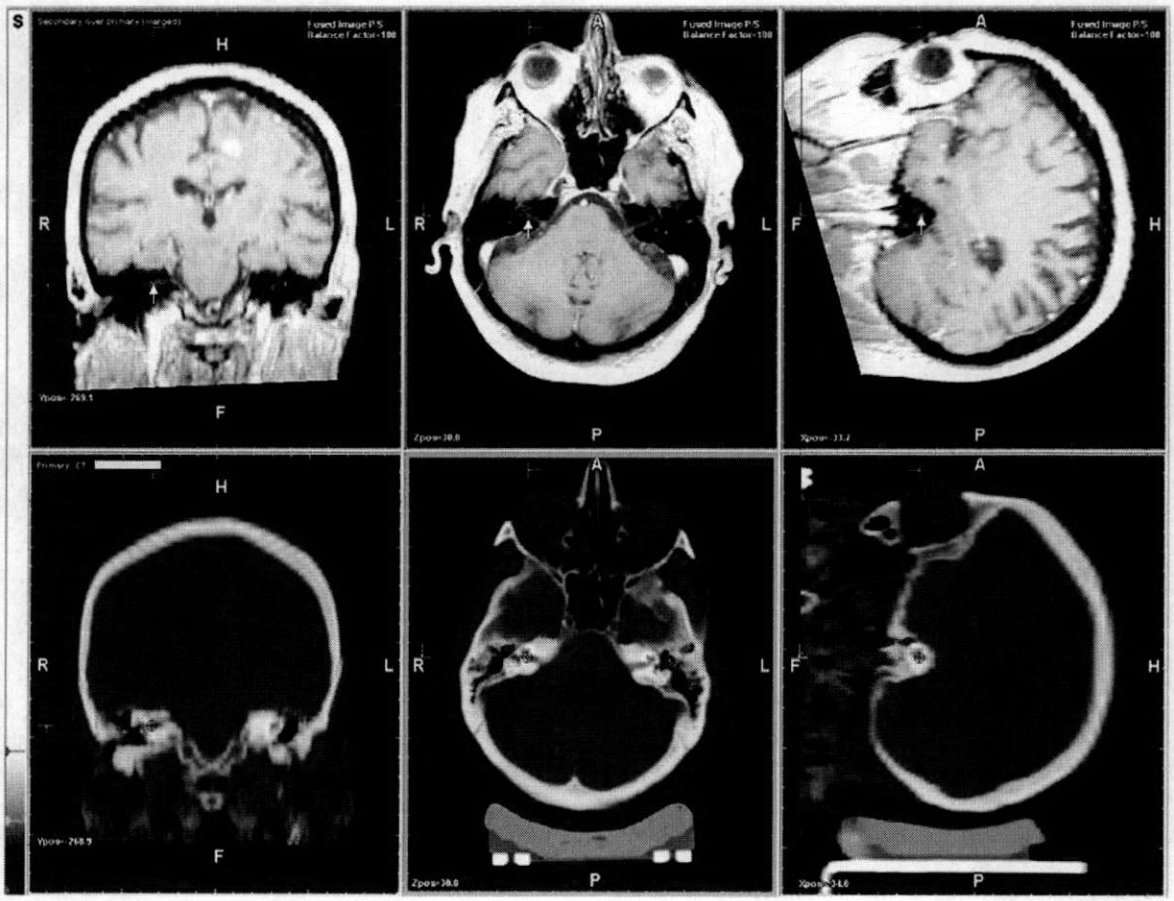


Fig. 4. Registered magnetic resonance (top row) and computed tomography (bottom row) brain scans. Coronal and sagittal reformats are left and right of the original transaxial scans. The cursor [circled cross (and arrow on MR)] is towards the distal end of the right vestibulocochlear nerve, an example of a suitable landmark (well localised in three-dimensions) for image registration verification.

parameters shows significant discrepancies between the methods then allows the selection of parameters to adjust in confirming that there has been an improvement, or in removing the degradation. A simple calculation of the geometrical combination of the full set of parameter differences, between methods, provides an indication of the overall worst spatial discrepancy within the relevant volume of the data. This may then be judged as to whether it is a significant discrepancy.

It is considered important that members of staff involved maintain their skill in performing manual registrations and in being able to detect errors in automated registrations, by performing a sufficient number of registrations (at least several per month).

Testing of registration software (for example at commissioning, or when changing to new protocols involving larger numbers of thinner slices, or different extents of inferior-superior coverage) may involve production of a series of MR scans that differ from each other by known shifts and rotations. Registration software can then be tested on its ability to identify correctly the known changes. It is generally not feasible to scan human subjects repeatedly for this purpose. Informative data can be obtained by performing such scans on test objects, but these do not present challenges to registration software that are identical to datasets from patients. Furthermore, obtaining sufficient time to acquire such multiple scans on a system as expensive as an MR scanner can be difficult. A useful alternative approach is to take a human DICOM dataset, computationally to apply known transformations (shifts and rotations or indeed scaling or localised distortion) to it, and to re-export the transformed dataset, still in DICOM format, for testing of the registration software [23]. For registration testing and QA, pseudo-anthropomorphic and anthropomorphic physical phantoms have been considered and investigated [24–26]. Processes of QA for registration are also required for serial acquisitions of three-dimensional image series [27].

Discussion and Conclusions

The ability of MRI to provide excellent depiction of the spatial relationships of normal and pathological soft tissues offers the potential of enabling more precise dose delivery and sparing of healthy tissues in radiotherapy planning (RTP).

MRI is less geometrically secure than CT, and it is important to be aware of geometrical distortion that may be present. Quality assurance should be carried out to establish the level of distortion within the FOVs that are used. For larger FOVs, distortion correction would be required, but is currently rarely used clinically, as commercial systems do not usually incorporate reliable functionality of this type. Significantly, when FOVs of around 23 cm or less are used, for many systems QA will reveal that distortion is small enough that correction is inessential for key MRI sequences.

Software for image registration of MR to CT has been available in commercial systems for many years, and recently the reliability of automated methods has improved, while faster computers can now provide the finished registration within a matter of seconds. There is considerable variation among systems in the functionality available for manual or interactive image registration. An important area of functionality to consider in evaluating systems is the provision of software tools such as linked cursors, to allow ergonomically efficient operator verification of registration quality, using anatomical reference points.

Developments in MRI scanner technology, such as wider bores and multi-element coils, are progressively offering advantages for RTP, in terms of the possibility of better matching of patient positioning versus CT in a greater range of anatomical regions, while allowing thinner slices for better image quality in orthogonal planes. MRI has always offered a variety of types of tissue contrast, according to the sequence and parameters deployed; it will become feasible to deploy additional MR sequences [28,29], further extending this variety. Just as for sequences commonly used currently for RTP, it will be necessary to put in place imaging protocols that have undergone appropriate QA and that are optimised for the requirements of RTP.

Acknowledgments

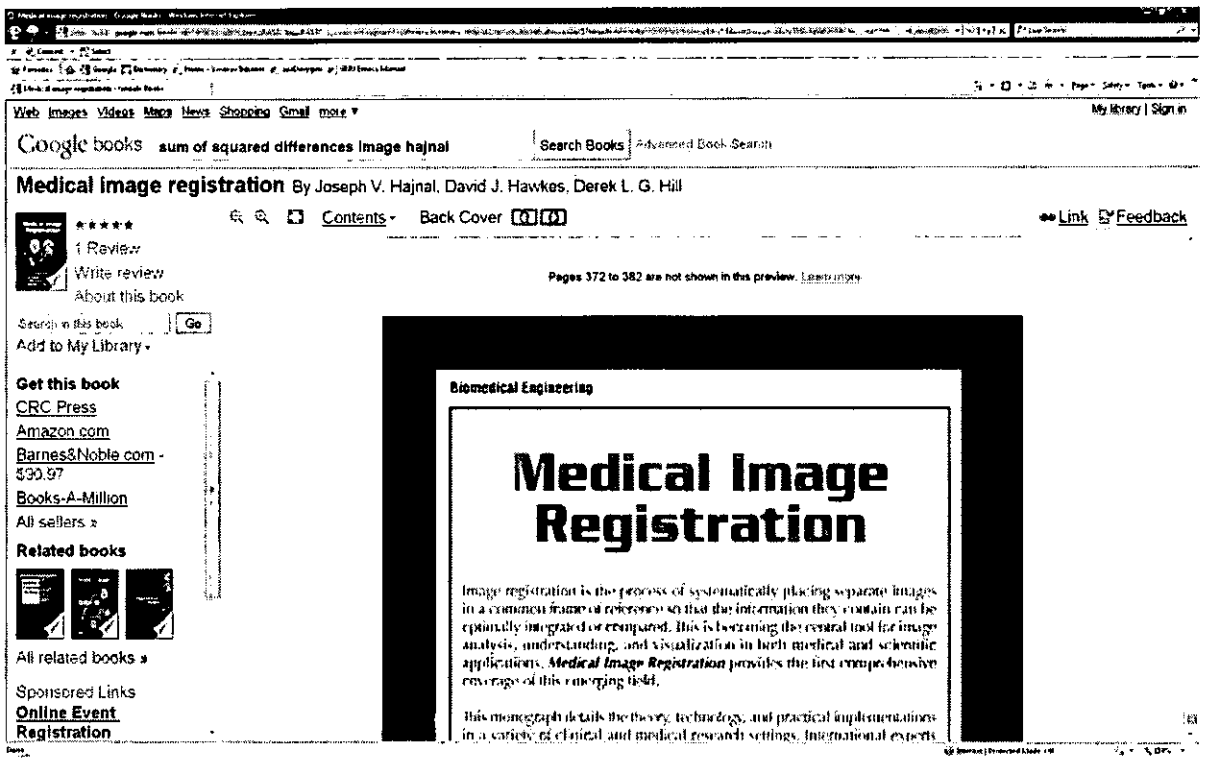
I would like to thank Dr Philip Mayles for encouraging my interest in this area of work, and Dr Christine Denby for bringing some recent literature to my attention. I am grateful for the beneficial involvement, in the implementation of the effective use of MR in RTP, of a large number of colleagues, including diagnostic radiographers, radiotherapy radiographers, clinical physicists, mechanical and electronic workshop and IT staff, diagnostic radiologists and clinical oncologists.

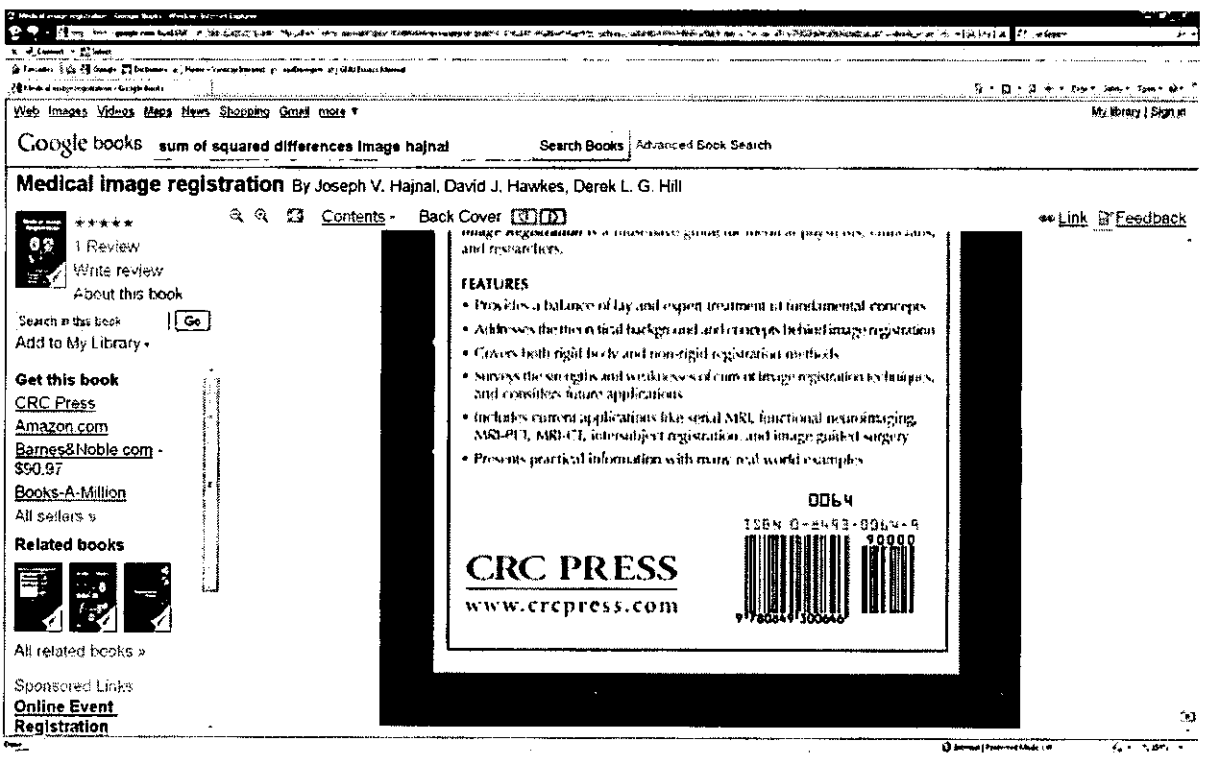
References

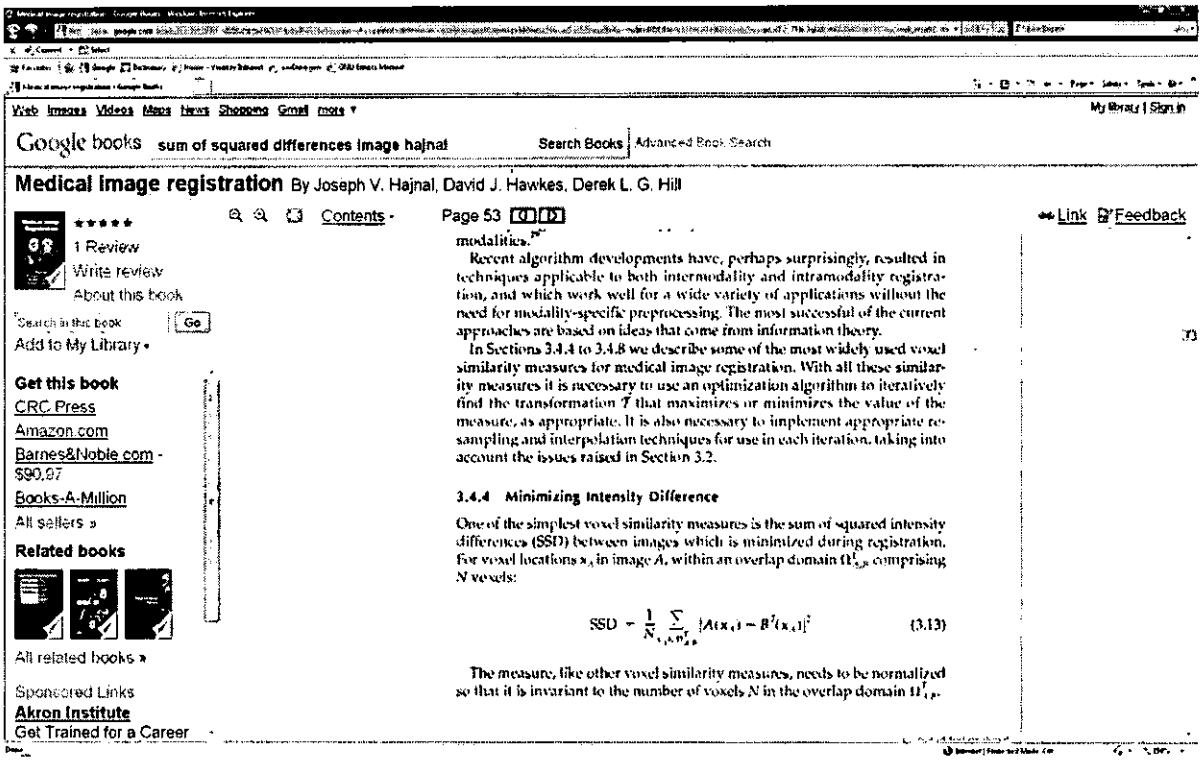
[1] Evans PM. Topical review: anatomical imaging for radiotherapy. *Phys Med Biol* 2008;53:R151–R191.
[2] Huq S, Mayles P, Besa de Carcer P, *et al. Transition from 2-D radiotherapy to 3-D conformal and intensity-modulated radiotherapy. IAEA-TECDOC-1588.* International Atomic Energy Agency; 2008.
[3] Beavis AW, Gibbs P, Dealey RA, Whitton VJ. Radiotherapy treatment planning of brain tumours using MRI alone. *Br J Radiol* 1998;71:544–548.
[4] Stanescu T, Jans H-S, Pervez N, Stavrev P, Fallone BG. A study on the magnetic resonance imaging (MRI)-based radiation treatment planning of intracranial lesions. *Phys Med Biol* 2008;53:3579–3593.
[5] Webster GJ, Kilgallon JE, Ho KF, Rowbottom CG, Slevin NJ, Mackay RI. A novel imaging technique for fusion of high-quality immobilised MR images of the head and neck with CT scans for radiotherapy target delineation. *Br J Radiol* 2009;82:497–503.
[6] Brunt JNH. A brief practical guide to using MRI, registered with CT, for radiotherapy planning. *Rad Magazine* 2005;31 (359):33–34.
[7] Department of Health. *Device bulletin. Safety guidelines for magnetic resonance imaging equipment in clinical use.* Department of Health; 2007.

[8] Lerski RA, De Wilde J, Boyce D, Ridgeway J. *Quality control in magnetic resonance imaging.* Report no. *80.* York: IPEM; 1999.
[9] Stanescu T, Jans H-S, Wachowicz K, Fallone BG. Investigation of a 3D system distortion correction method for MR images. *J Appl Clin Med Phys* 2010;11(1):200–216.
[10] Doran SJ, Charles-Edwards EM, Reinsberg SA, Leach MO. A complete distortion correction for MR images: I. Gradient warp correction. *Phys Med Biol* 2005;50:1343–1361.
[11] Reinsberg SA, Doran SJ, Charles-Edwards EM, Leach MO. A complete distortion correction for MR images: II. Rectification of static-field inhomogeneities by similarity-based profile mapping. *Phys Med Biol* 2005;50:2651–2661.
[12] Raaymakers BW, Lagendijk JJW, Overweg J, *et al.* Integrating a 1.5 T MRI scanner with a 6 MV accelerator: proof of concept. *Phys Med Biol* 2009;54:N229–N237.
[13] Fallone BG, Murray B, Rathee S, *et al.* First MR images obtained during megavoltage photon irradiation from a prototype integrated linac-MR system. *Med Phys* 2009;36(6):2084–2088.
[14] Lamey M, Burke B, Blosser E, Rathee S, De Zanche N, Fallone BG. Radio frequency shielding for a linac-MRI system. *Phys Med Biol* 2010;55:995–1006.
[15] ViewRay. Available at: http://www.viewray.com/technology.php; (accessed 20 April 2010).
[16] Maintz JBA, Viergever MA. A survey of medical image registration. *Med Image Anal* 1998;2(1):1–36.
[17] Hill DL, Batchelor PG, Holden M, Hawkes DJ. Medical image registration. *Phys Med Biol* 2001;46(3):R1–R45.
[18] Veninga T, Huisman H, van der Maazen RW, Huizenga H. Clinical validation of the normalized mutual information method for registration of CT and MR images in radiotherapy of brain tumors. *J Appl Clin Med Phys* 2004;5(3):66–79.
[19] Hill DL, Hawkes DJ, Gleeson MJ, *et al.* Accurate frameless registration of MR and CT images of the head: applications in planning surgery and radiation therapy. *Radiology* 1994;191:447–454.
[20] Brock KK, Deformable Registration Accuracy Consortium. Results of a multi-institution deformable registration accuracy study (MIDRAS). *Int J Radiat Oncol Biol Phys* 2010;76:583–596.
[21] van Beek S, van Kranen S, Mencarelli A, *et al.* First clinical experience with a multiple region of interest registration and correction method in radiotherapy of head-and-neck cancer patients. *Radiother Oncol* 2010;94:213–217.
[22] Kessler ML. Image registration and data fusion in radiation therapy. *Br J Radiol* 2006;79:S99–S108.
[23] Oncology Systems Limited. ImSimQA, virtual phantoms for real QA. 2008 IM-SimQAJULY08.pdf. Available at: www.osl.uk.com; (accessed 20 April 2010).
[24] Moore CS, Liney GP, Beavis AW. Quality assurance of registration of CT and MRI data sets for treatment planning of radiotherapy for head and neck cancers. *J Appl Clin Med Phys* 2004;5:25–35.
[25] Shmueli K, Thomas DL, Ordidge RJ. Design, construction and evaluation of an anthropomorphic head phantom with realistic susceptibility artifacts. *J Magn Reson Imaging* 2007;26:202–207.
[26] Mutic S, Dempsey JF, Bosch WR, *et al.* Multimodality image registration quality assurance for conformal three-dimensional treatment planning. *Int J Radiat Oncol Biol Phys* 2001;51:255–260.
[27] Sharpe M, Brock KK. Quality assurance of serial 3d image registration, fusion, and segmentation. *Int J Radiat Oncol Biol Phys* 2008;71:S33–S37.
[28] Khoo VS, Joon DL. New developments in MRI for target volume delineation in radiotherapy. *Br J Radiol* 2006;79:S2–S15.
[29] Payne GS, Leach MO. Applications of magnetic resonance spectroscopy in radiotherapy treatment planning. *Br J Radiol* 2006;79:S16–S26.

Medical image registration, J Hajnal, D Hawkes, D Hill, Page 53.







Web Images Videos Maps News Shopping Gmail more ▾ My Library | Sign in

Google books sum of squared differences image hajnal Search Books Advanced Book Search

Medical image registration By Joseph V. Hajnal, David J. Hawkes, Derek L. G. Hill

Contents - Page 53 Link Feedback

★★★★★
1 Review
Write review
About this book

Search in the book Go
Add to My Library ▾

Get this book
CRC Press
Amazon.com
Barnes&Noble.com - $90.97
Books-A-Million
All sellers »

Related books

All related books »

Sponsored Links
Akron Institute
Get Trained for a Career

modalities.[87]

Recent algorithm developments have, perhaps surprisingly, resulted in techniques applicable to both intermodality and intramodality registration, and which work well for a wide variety of applications without the need for modality-specific preprocessing. The most successful of the current approaches are based on ideas that come from information theory.

In Sections 3.4.4 to 3.4.8 we describe some of the most widely used voxel similarity measures for medical image registration. With all these similarity measures it is necessary to use an optimization algorithm to iteratively find the transformation T that maximizes or minimizes the value of the measure, as appropriate. It is also necessary to implement appropriate resampling and interpolation techniques for use in each iteration, taking into account the issues raised in Section 3.2.

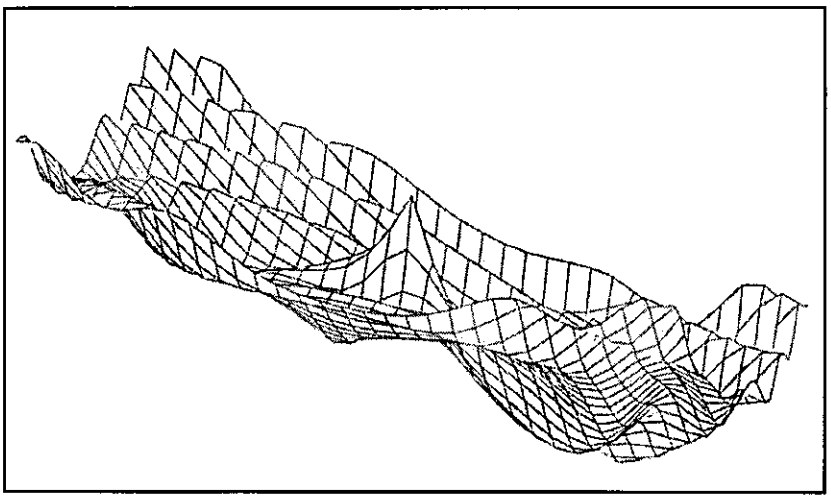
3.4.4 Minimizing Intensity Difference

One of the simplest voxel similarity measures is the sum of squared intensity differences (SSD) between images which is minimized during registration. For voxel locations $\mathbf{x}_i$ in image A, within an overlap domain $\Omega'_{A,B}$ comprising N voxels:

$$\text{SSD} = \frac{1}{N} \sum_{\mathbf{x}_i \in \Omega'_{A,B}} [A(\mathbf{x}_i) - B'(\mathbf{x}_i)]^2 \qquad (3.13)$$

The measure, like other voxel similarity measures, needs to be normalized so that it is invariant to the number of voxels N in the overlap domain $\Omega'_{A,B}$.

Iterative Methods for Optimization



C. T. Kelley
North Carolina State University
Raleigh, North Carolina

siam
Society for Industrial and Applied Mathematics
Philadelphia

Chapter 3

Global Convergence

The locally convergent algorithms discussed in Chapter 2 can and do fail when the initial iterate is not near the root. The reasons for this failure, as we explain below, are that the Newton direction may fail to be a direction of descent for f and that even when a search direction is a direction of decrease of f, as $-\nabla f$ is, the length of the step can be too long. Hence, taking a Newton (or Gauss–Newton, or inexact Newton) step can lead to an increase in the function and divergence of the iteration (see exercise 3.5.14 for two dramatic examples of this). The *globally convergent* algorithms developed in this chapter partially address this problem by either finding a local minimum or failing in one of a small number of easily detectable ways.

These are not algorithms for global optimization. When these algorithms are applied to problems with many local minima, the results of the iteration may depend in complex ways on the initial iterate.

3.1 The Method of Steepest Descent

The *steepest descent direction* from x is $d = -\nabla f(x)$. The *method of steepest descent* [52] updates the current iteration x_c by the formula

$$x_+ = x_c - \lambda \nabla f(x_c). \tag{3.1}$$

If we take the simple choice $\lambda = 1$, then x_+ is not guaranteed to be nearer a solution than x_c, even if x_c is very near a solution that satisfies the standard assumptions. The reason for this is that, unlike the Newton direction, the steepest descent direction scales with f. The Newton step, on the other hand, is the same for f as it is for cf for any $c \neq 0$ but need not be a direction of decrease for f.

To make the method of steepest descent succeed, it is important to choose the *steplength* λ. One way to do this, which we analyze in §3.2, is to let $\lambda = \beta^m$, where $\beta \in (0, 1)$ and $m \geq 0$ is the smallest nonnegative integer such that there is *sufficient decrease* in f. In the context of the steepest descent algorithm, this means that

$$f(x_c - \lambda \nabla f(x_c)) - f(x_c) < -\alpha \lambda \|\nabla f(x_c)\|^2. \tag{3.2}$$

This strategy, introduced in [7] and called the *Armijo rule*, is an example of a *line search* in which one searches on a ray from x_c in a direction in which f is locally decreasing. In (3.2), $\alpha \in (0, 1)$ is a parameter, which we discuss after we fully specify the algorithm. This strategy of repeatedly testing for sufficient decrease and reducing the stepsize if the test is failed is called *backtracking* for obvious reasons.

The motivation for (3.2) is that if we approximate f by the *linear model*

$$m_c = f(x_c) + \nabla f(x_c)(x - x_c),$$

then the reduction in the model (i.e., the *predicted reduction* in f) is

$$pred = m_c(x_c) - m_c(x_+) = \lambda \|\nabla f(x_c)\|^2.$$

(3.2) says that the *actual reduction* in f

$$ared = f(x_c) - f(x_+)$$

is at least as much as a fraction of the predicted reduction in the linear model. The parameter α is typically set to 10^{-4}.

The reason we demand sufficient decrease instead of *simple decrease* (i.e., $f(x_c) < f(x_+)$ or $\alpha = 0$) is largely theoretical; a nonzero value of α is required within the proof to insure that the iteration does not stagnate before convergence.

ALGORITHM 3.1.1. `steep`$(x, f, kmax)$

1. *For* $k = 1, \ldots, kmax$
   (a) *Compute* f *and* ∇f*; test for termination.*
   (b) *Find the least integer* $m \geq 0$ *such that* (3.2) *holds for* $\lambda = \beta^m$.
   (c) $x = x + \lambda d$.
2. *If* $k = kmax$ *and the termination test is failed, signal failure.*

The termination criterion could be based on (2.12), for example.

3.2 Line Search Methods and the Armijo Rule

We introduce a few new concepts so that our proof of convergence of Algorithm `steep` will also apply to a significantly more general class of algorithms.

DEFINITION 3.2.1. *A vector* $d \in R^N$ *is a descent direction for* f *at* x *if*

$$\left.\frac{df(x + td)}{dt}\right|_{t=0} = \nabla f(x)^T d < 0.$$

Clearly the steepest descent direction $d = -\nabla f(x)$ is a descent direction. A *line search algorithm* searches for decrease in f in a descent direction, using the Armijo rule for stepsize control, unless $\nabla f(x) = 0$.

We will consider descent directions based on *quadratic models* of f of the form

$$m(x) = f(x_c) + \nabla f(x_c)^T(x - x_c) + \frac{1}{2}(x - x_c)^T H_c(x - x_c),$$

where H_c, which is sometimes called the *model Hessian*, is spd. We let $d = x - x_c$ be such that $m(x)$ is minimized. Hence,

$$\nabla m(x) = \nabla f(x_c) + H_c(x - x_c) = 0$$

and hence

$$d = -H_c^{-1} \nabla f(x_c). \tag{3.3}$$

Fast free-form deformable registration via calculus of variations

2004 Phys. Med. Biol. 49 3067

(http://iopscience.iop.org/0031-9155/49/14/003)

INSTITUTE OF PHYSICS PUBLISHING PHYSICS IN MEDICINE AND BIOLOGY

Phys. Med. Biol. **49** (2004) 3067–3087 PII: S0031-9155(04)75850-9

Fast free-form deformable registration via calculus of variations

Weiguo Lu[1], **Ming-Li Chen**[2], **Gustavo H Olivera**[1,2], **Kenneth J Ruchala**[1] **and Thomas R Mackie**[1,2]

[1] TomoTherapy Inc., 1240 Deming Way, Madison, WI 53717, USA
[2] University of Wisconsin-Madison, 1300 University Avenue Madison, WI 53705, USA

E-mail: wlu@tomotherapy.com

Received 5 February 2004
Published 28 June 2004
Online at stacks.iop.org/PMB/49/3067
doi:10.1088/0031-9155/49/14/003

Abstract

In this paper, we present a fully automatic, fast and accurate deformable registration technique. This technique deals with free-form deformation. It minimizes an energy functional that combines both similarity and smoothness measures. By using calculus of variations, the minimization problem was represented as a set of nonlinear elliptic partial differential equations (PDEs). A Gauss–Seidel finite difference scheme is used to iteratively solve the PDE. The registration is refined by a multi-resolution approach. The whole process is fully automatic. It takes less than 3 min to register two three-dimensional (3D) image sets of size 256 × 256 × 61 using a single 933 MHz personal computer. Extensive experiments are presented. These experiments include simulations, phantom studies and clinical image studies. Experimental results show that our model and algorithm are suited for registration of temporal images of a deformable body. The registration of inspiration and expiration phases of the lung images shows that the method is able to deal with large deformations. When applied to the daily CT images of a prostate patient, the results show that registration based on iterative refinement of displacement field is appropriate to describe the local deformations in the prostate and the rectum. Similarity measures improved significantly after the registration. The target application of this paper is for radiotherapy treatment planning and evaluation that incorporates internal organ deformation throughout the course of radiation therapy. The registration method could also be equally applied in diagnostic radiology.

0031-9155/04/143067+21$30.00

1. Introduction

Medical image registration has been the subject of extensive study in the literature. Recent reviews have appeared in several journal publications (Brown 1992, van den Elsen *et al* 1993, Maintz and Viergever 1998, Hill *et al* 2001, Zitova and Flusser 2003) and books (Hajnal *et al* 2001, Gee *et al* 2003). Image registration computes a transformation that relates the points in one image to their corresponding points in the other. The transformation can be global or local, rigid or deformable (Lu 2001). The rigid-body model simplifies the registration process, but rigid-body transforms have limited applicability because many organs deform substantially. Rigid-body is a good model for individual bone motion. For example, by far the most important part of the body registered in this way is the cranium and its contents—the brain. However, the extension of the rigid-body model to extra-cranial targets is more difficult, because the assumption of a rigid-body is not valid within the abdomen, pelvis, thorax or extremities. For other regions of the body in the vicinity of bone, the rigid-body model may be used, but may cause large error. Internal structures do enlarge and contract due to normal physiological functions, the region of interest volume does not necessarily maintain a fixed relationship to bony structures. For most organs in the body, many more degrees of freedom (DOF) are needed to describe the different deformations with adequate accuracy.

There have been many approaches to deformable registration. As for the similarity measures (Rogelj and Kovacic 2001) considered, most techniques could be categorized as point-based and voxel-based methods. Point-based techniques minimize the distance between features such as points, curves or surfaces of corresponding anatomical structures. They require the identification and matching of these features on both reference and test images. This process usually needs certain amount of human interactions. After point matching, the remaining procedure of registration is only interpolation or approximation. Voxel-based methods use similarity measures such as sum of squared distance, cross correlation or mutual information (Maes *et al* 1997, Studholme *et al* 1999, Pluim *et al* 2003) between images. They have the advantage that they do not require any feature extraction process.

As for the spatial transformations, the deformations can be grouped as polynomial transformation (Woods *et al* 1998), spline and thin-plate spline (Bookstein 1989) transformation and free-form deformation (FFD) (Sederberg and Parry 1986). The DOF of polynomial transformation is defined by its order. Polynomial transformations cannot accommodate local shape changes but only model global shape changes. Therefore, their ability to recover anatomical shape variability is quite limited. In addition, the degrees of higher order polynomials tend to introduce artefacts such as oscillations. Spline and thin-plate spline transformation requires certain number of control points. The DOF of spline and thin-plate spline transformation is proportional to the number of control points used. Each control point has global influence on the transformation. In many cases, this global influence of control points is undesirable since it becomes difficult to model local deformations. In addition, the computation complexity makes using large number of control points prohibitive. Usually, an FFD is the best choice for many situations for its ability to model local deformation, its unlimited DOF, and its computation efficiency.

The FFD must be constrained/regularized because of ill-posedness of the problem (the existence of many possible solutions). Physical models, such as linear elastic models (Burr 1981, Bajcsy and Kovacic 1989, Lu *et al* 2001, Lu *et al* 2001), viscoelastic models (Miller *et al* 1993), viscous fluid models (Christensen *et al* 1996, Wang and Staib 1998) and biomechanical model (Hagemann *et al* 1999, Kyriacou and Davatzikos 1999, Kyriacou *et al* 1999) are widely used to enforce topological properties of the deformation and constrain the

enormous solution spaces. Generally, the whole image is modelled as a physical body, and the similarity measures between two images act as external forces that 'stretch' the body. These external forces are counterbalanced by physical constraints. The ultimate goal is the determination of a state with minimum energy whose resulting deformation defines the registration. The problems associated with finding the minimum energy state or equilibrium usually involve iterative numerical methods. Physical model based methods usually have heavy computation demands, and even require massive parallel computers (Christensen *et al* 1996). FFD can also be modelled as optical flow (Horn and Schunck 1981) or 'demons' (Thirion 1995, Thirion 1998). In these cases, regularization is applied by filtering/smoothing the deformation field (Thirion 1998, Guimond *et al* 2001).

Conventionally, in 3D radiation therapy, a planning image is assumed to be a valid representation of the patient throughout the course of radiotherapy. Translation and perhaps even rotation set-up verification can be used to perform 'rigid-body' realignments of the patient based on transmission planar images. Set-up verification (or rigid-body registration) is not adequate for modern radiotherapy applications. Internal organ deformation throughout the course of radiation therapy is problematic for accurately evaluating the cumulative dose delivered (Yan *et al* 1999, Lu *et al* 2000, Olivera *et al* 2003). Deformable registration is a technique to account for internal organ deformation. Deformable registration is also the basis to build four-dimensional (4D) model for radiation therapy planning and evaluation (Brock *et al* 2003, Zhang *et al* 2004). To apply deformable registration routinely in radiation therapy, a desirable method should be accurate, fast and fully automatic. In addition, it should have enough number of degrees of freedom to cope with large deformations. The application nature excludes the point-based methods that require manually selecting a large number of landmarks and physical model based methods that require prior knowledge of the material properties and heavy computation demands.

In this paper, we present a new, fast and accurate voxel similarity based free-form deformable registration method. The deformable registration problem is modelled as a functional minimization problem. Rather than using a complicated physics model, we use the smoothness of displacement field as a constraint. Therefore, no prior knowledge about the physical properties of the patient is required. The smoothness constraints are somehow similar as Thirion's 'demons' method (Thirion 1995, 1998, Ibanez *et al* 2003). In Thirion's method, the smoothness constraints are explicit in the algorithm since the displacement fields are convolved with a Gaussian filter after each iteration. In our approach, this constraint is implicitly built into the objective functional, therefore, the optimization process itself addresses the trade-off between the similarity measures and smooth constraints. We use calculus of variations to represent the optimization problem as a set of nonlinear elliptic partial differential equations (PDEs). The PDEs are iteratively solved using a Gauss–Seidel finite difference scheme. A multi-resolution approach is used to refine the registration. The method and our implementation are general and can be used to register any dimension data, such as 1D signals, 2D images or 3D volume data. The whole process is fully automatic and efficient. The registration of two 3D volumes is computed within minutes using a single personal computer. Theoretically, the method itself does not eliminate local minima. But in practice, local minima are rarely met due to the combination of multi-resolution technique and Gause–Seidel updating scheme. With the combination of all these techniques, images with large deformations can also be accurately registered.

The target application of this paper is radiotherapy planning and evaluation that incorporates internal organ deformation. Multiple CT scans of the same patient are required. But the method itself could also be equally applicable to diagnostic radiology, both CT and MR in situations where patients are scanned on multiple occasions.

2. Theory

2.1. Free-form deformation

When the relative positions of particles in a continuous body are altered, we say that the body is strained and the change in the relative position of points is a deformation. To describe the deformation of the body, we need to know the position of any point in the body with respect to a reference configuration. We use a Eulerian reference frame. The Eulerian description of material deformations specifies the time evolution of particle positions and velocities as observed at fixed points. At the reference state, the location of material point P is $\boldsymbol{x} = (x^1, x^2, x^3)$. We use $\boldsymbol{x}$ as a label for that point. As time goes on, the point moves. Its location has history: $\boldsymbol{x}' = \boldsymbol{x}'(\boldsymbol{x}, t)$ refers to the same reference frame. Let us define a displacement mapping from a domain A(reference) to domain B(test):

$$\boldsymbol{u} : A \rightarrow B \qquad \boldsymbol{x}'(\boldsymbol{x}, t) = \boldsymbol{x} + \boldsymbol{u}(\boldsymbol{x}, t). \tag{2.1}$$

Equation (2.1) means that a point originating at $\boldsymbol{x}$ will move to point $\boldsymbol{x} + \boldsymbol{u}(\boldsymbol{x}, t)$ at time t. We call (2.1) a registration map. The vectors $\boldsymbol{u}(\boldsymbol{x}, t)$ are the displacement vectors. The objective of deformable registration is to find out these displacement vectors.

The deformation can be local or global, small or large, elastic or inelastic. The most general deformation is 'free-form deformation'; that is, each voxel can move freely without any explicit constrains. In free-form deformation, the degrees of freedom are $3N$, where N is the total number of voxels.

2.2. Minimization of the energy functional

In this paper, we describe the free-form deformable registration problem as finding the displacement fields $\hat{\boldsymbol{u}}$ that minimize the energy functional $\varepsilon(\boldsymbol{u})$:

$$\hat{\boldsymbol{u}} = \arg \min_{\boldsymbol{u}} \varepsilon(\boldsymbol{u}) \tag{2.2}$$

where

$$\varepsilon(\boldsymbol{u}) = \int_{\boldsymbol{x} \in R^3} \left[R^2(\boldsymbol{x}, \boldsymbol{u}) + \lambda \sum_{i=1}^{3} \sum_{j=1}^{3} \left(v_j{}^i \right)^2 \right] \mathrm{d}\boldsymbol{x}. \tag{2.3}$$

Here $R(\boldsymbol{x}, \boldsymbol{u}) = B(\boldsymbol{x} + \boldsymbol{u}) - A(\boldsymbol{x})$ is the residual. A is the reference data and B is the test data, and $v_j{}^i \equiv \frac{\partial u^i}{\partial x^j}$. Here we assume that two images are similar in intensity level so that intensity residual is appropriate to describe their misalignment.

This variational formulation follows a standard principle of making the result smooth when there are no data. In particular, we see that when the residual square R^2 is large, the first term dominates, and is minimized by setting $R = 0$, that is, two images match each other in voxel intensity. On the other hand, when R^2 is small, the energy is dominated by summation of the squares of the partial derivatives of the displacement vector, yielding a slowly varying field. Equation (2.2) is a nonlinear minimization problem with thousands to millions of variables. It is prohibitive to solve it with the conventional optimization algorithm on a personal computer within reasonable time.

2.3. Calculus of variations

By using the calculus of variations (Keener 1988, Xu 2000), it can be shown that the displacement field can be found by solving the following Euler–Lagarange equations,

$$\lambda \nabla^2 \boldsymbol{u} - R(\boldsymbol{x}, \boldsymbol{u}) \frac{\partial R(\boldsymbol{x}, \boldsymbol{u})}{\partial \boldsymbol{u}} = 0 \tag{2.4}$$

where ∇^2 is the Laplacian operator with respect to x; $\nabla^2 = \sum_{i=1}^{3} \frac{\partial^2}{\partial x^{i2}}$ and λ is a constant weight.

Equation (2.4) provides further intuition behind the displacement formulation. We note that the region where two images have matched intensity ($R = 0$), the second term in equation (2.4) is zero. Therefore, within such a region, displacement vector u is determined by Laplacian equation, and the resulting displacement vector is interpolated from other unmatched region. This results in a smoothly varying displacement field.

As in equation (2.4), the deformable registration problem becomes the problem of solving the nonlinear elliptic partial differential equations (PDE).

Further expanding $\frac{\partial R(x,u)}{\partial u}$, we have

$$\begin{aligned}\frac{\partial R(x,u)}{\partial u} &= \frac{\partial B(x+u)}{\partial u} \\ &= \left.\frac{\partial B(z)}{\partial z}\right|_{z=x+u} \frac{\partial z}{\partial u} \\ &= \left.\frac{\partial B(z)}{\partial z}\right|_{z=x+u} \\ &= (\nabla B)(x+u) \\ &= g(x+u)\end{aligned} \tag{2.5}$$

where $g(x) = \nabla B(x)$.

So we have

$$\lambda\nabla^2 u - [B(x+u) - A(x)]g(x+u) = 0. \tag{2.6}$$

3. Numerical implementation

3.1. Finite difference scheme

Equation (2.6) is a system of nonlinear elliptic PDE. We use a finite difference scheme to solve it.

To set up the finite difference scheme, the problem needs to be discretized. Let indices i, j, k correspond to x, y and z coordinates, ($i = 1, 2, \ldots, L_x$, $j = 1, 2, \ldots, L_y$, $k = 1, 2, \ldots, L_z$). Let indices n correspond to dimensions ($n = 1, 2, 3$). Let the combined indices (i, j, k) be organized as in a certain sequence (for example, a 3D volume is organized as a 1D array), and let the indices for that parametric sequence be m ($m = 1, 2, \ldots, N$), N is the total number of voxels.

Discretizing the left-hand side of equation (2.6), and let it be $L_{m,n}$, we have

$$L_{m,n} = \lambda\nabla^2(u_{m,n}) - [B^*{}_m - A_m]g^*{}_{m,n} \tag{3.1}$$

Tri-linear interpolation is used to calculate transformed tested image $B^*{}_m$ and its gradient $g^*{}_{m,n}$,

$$B^*{}_m = B(x_m + u_m) \tag{3.2}$$

$$g^*{}_{m,n} = g_n(x_m + u_m) \tag{3.3}$$

and

$$\nabla^2(u_{m,n}) = \frac{u_{i+1,j,k,n} + u_{i-1,j,k,n} + u_{i,j+1,k,n} + u_{i,j-1,k,n} + u_{i,j,k+1,n} + u_{i,j,k-1,n}}{6} - u_{i,j,k,n} \tag{3.4}$$

which is the difference between the displacement of 6 neighbours about a point and the point itself. Note that in equation (3.4), the index m on the left-hand side is decomposed as indexes i, j, k on the right-hand side.

One step of Newton iteration was used to update $u_{m,n}$, that is,

$$u^{\text{new}}{}_{m,n} = u^{\text{old}}{}_{m,n} - \frac{L^{\text{old}}{}_{m,n}}{\partial L^{\text{old}}{}_{m,n}/\partial u_{m,n}} \tag{3.5}$$

where

$$\frac{\partial L_{m,n}}{\partial u_{m,n}} = \lambda \frac{\partial(\nabla^2(u_{m,n}))}{\partial u_{m,n}} - \left(\frac{\partial(B^*{}_m)}{\partial u_{m,n}} g^*{}_{m,n} + \frac{\partial(g^*{}_{m,n})}{\partial u_{m,n}} (B^*{}_m - A_m) \right). \tag{3.6}$$

From equation (3.3), we have

$$\frac{\partial(\nabla^2(u_{m,n}))}{\partial u_{m,n}} = -1 \tag{3.7}$$

and from equation (2.5), we have

$$\frac{\partial B^*{}_m}{\partial u_{m,n}} = g^*{}_{m,n}. \tag{3.8}$$

We discard the second derivative term $\partial g^*{}_{m,n}/\partial u_{m,n}$, then we have

$$\frac{\partial L_{m,n}}{\partial u_{m,n}} = -(\lambda + (g^*{}_{m,n})^2). \tag{3.9}$$

Therefore, our update scheme is

$$u^{\text{new}}{}_{m,n} = u^{\text{old}}{}_{m,n} + \frac{L^{\text{old}}{}_{m,n}}{\lambda + \left(g^{*\text{old}}{}_{m,n}\right)^2}. \tag{3.10}$$

Gauss–Seidel method (Press *et al* 1992) is used to calculate $L^{\text{old}}{}_{m,n}$ and $g^{*\text{old}}{}_{m,n}$, that is, we use the updated version of $u_{m,n}$ as soon as they become available.

In summary, the update scheme for each iteration is as follows:

For $n = 1, 2, 3$

  For $m = 1, 2, \ldots, N$

    Calculate $\nabla^2(u_{m,n})$ according to (3.4)

    Calculate $B^*{}_m$ and $g^*{}_{m,n}$ using tri-linear interpolation (3.2) and (3.3)

    Calculate $L_{m,n}$ according to (3.1)

    Update $u_{m,n}$ according to (3.10)

  Endfor

Endfor.

3.2. Multi-resolution strategy

Multi-resolution technique has emerged as a standard technique for deformable registration (Bajcsy and Kovacic 1989, Lester and Arridge 1999, Hellier *et al* 2001, Mattes *et al* 2003). Multi-resolution technique is a systematic way for structuring local information into global. The major motivation for using a multi-resolution technique is to alleviate the deficiencies of the one-resolution model (Bajcsy and Kovacic 1989), e.g., the limitation of small deformations, the limitation to propagate the boundary deformation further into the inner region and the computational complexity.

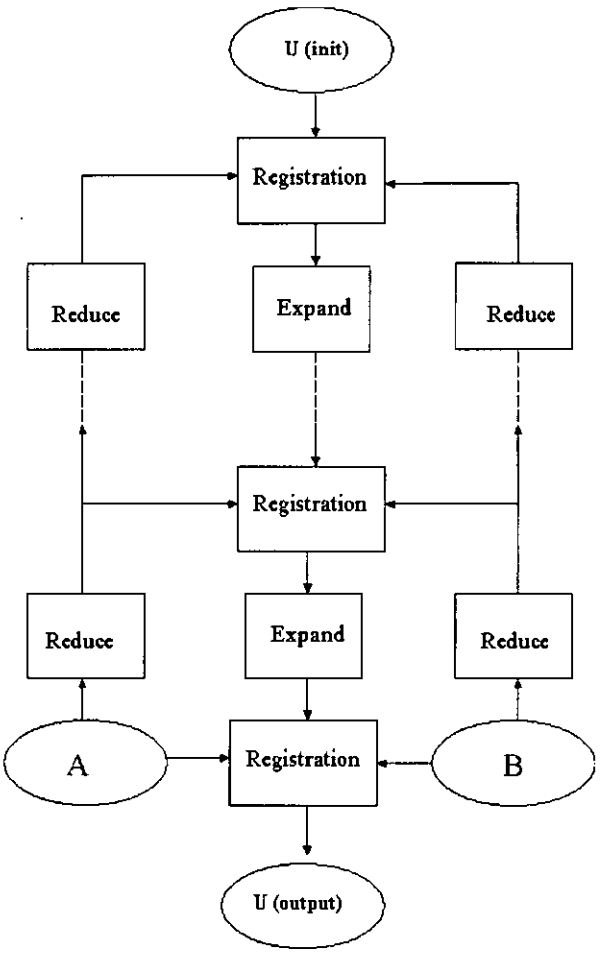


Figure 1. Flowchart for multi-resolution registration. A and B are reference and test images, respectively. Two image pyramids are constructed before the registration (left-most and right-most column). U are displacement fields. The algorithm accepts an initial displacement, which could be from any initial guess or previous registration. The registration is from top (coarsest) to bottom (finest). Each level registration is used as the initial guess for the latter level. The reduce operation means down-sampling of the images. The expand operation means up-sampling and scaling of the displacement field.

In multi-resolution registration, a coarse level registration is obtained first, which is later improved with the finer resolution calculation.

For the multi-resolution algorithm (figure 1), an image pyramid is constructed. The bottom of the pyramid is the original image. Each level of pyramid is the down-sampled version (*reduce* operation) of its lower level. The down-sample factor is 2 when the down-sampled dimension is not less than 32 and 1 otherwise. As for the optimization iterations, each level uses twice the number of iterations than that of its lower level. For example, suppose the original image set has dimension $512 \times 256 \times 68$ and bottom level uses 8 iterations, then the pyramid has 5 levels, each level has dimensions, from bottom to top, $(512 \times 256 \times 68)$, $(256 \times 128 \times 34)$, $(128 \times 64 \times 34)$, $(64 \times 32 \times 34)$ and $(32 \times 32 \times 34)$ respectively. The

number of iterations for each level is 8, 16, 32, 64 and 128, respectively. Registration starts on the top level (with coarsest resolution). The resulting displacement field u from the coarse level, after it is appropriately expanded (*expand* operation), is used as the initial displacement field for the finer level registration.

4. Evaluation

There are generally two criteria to evaluate a deformable registration algorithm:

(a) The difference between the calculated displacement field with that of the gold standard.
(b) The similarity measures between the deformed test image and the reference image.

The first criterion requires a gold standard displacement field, which only exists on the simulation study or well-designed phantom study. The second criterion could apply on any image data sets, but it cannot tell how well the registration matches the real deformation.

As for the similarity measures, we use (normalized) cross correlation (CC) and mutual information (MI). The CC is defined as

$$\mathrm{CC} = \frac{\sum (A(\boldsymbol{x}) - \bar{A})(B(\boldsymbol{x}) - \bar{B})}{\left(\sum (A(\boldsymbol{x}) - \bar{A})^2 \sum (B(\boldsymbol{x}) - \bar{B})^2\right)^{\frac{1}{2}}}. \tag{4.1}$$

Here, $\bar{A}$ and $\bar{B}$ are the mean intensity of the reference and the test image, respectively. If the reference and the target image are identical (perfect registration), CC is unity.

The MI is defined as (Studholme *et al* 1999)

$$\mathrm{MI} = -\sum p_{AB}(i, j) \log \frac{p_{AB}(i, j)}{p_A(i) p_B(j)} \tag{4.2}$$

where i, j are the intensity level of the reference image and the target image, respectively. $p(\cdot)$ is the probability distribution. The larger the MI, the better the registration.

5. Experiments and results

The presented algorithm could be applied to any dimension (e.g., a 1D signal, a 2D slice, and a 3D volume data) image. The algorithm is implemented in C++ programming language. The same code is used to register any dimension image data. There are only two parameters needed to run the code. One is the Laplacian weight λ, the other is the number of iterations for the finest resolution. Fortunately, as we have extensively tested, the algorithm is not very sensitive to these two parameters, and a rough guess is good enough. In the following tests, we always use $\lambda = 0.1$, and use 16 iterations for the finest level.

5.1. Synthetic data

1D registration. Figure 2 illustrates how the algorithm works for 1D signal. Here we have reference signal (red solid line) and test signal (green dashed line). The objective of deformable registration is to deform the test signal so that it approaches the reference signal. The blue dashed line shows the deformed test signal after each iteration. The bottom panel shows the growth of displacement field after each iteration.

2D registration. In this experiment, we use a 2D image, which is CT slice (512 × 512), as the reference image. This reference image is deformed using a known formula, which we call harmonic deformation. The harmonic deformation is defined as

$$x'(x, y) = (1 + b\cos m\theta)x \qquad y'(x, y) = (1 + b\cos m\theta)y. \tag{5.1}$$

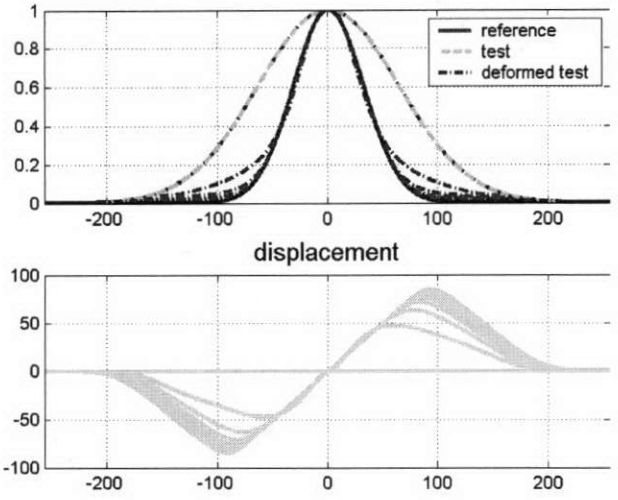


Figure 2. Illustration of 1D signal registration. The registration objective is to deform the test signal to approach the reference signal. The top panel shows the reference (red solid line), test (green dasheded line) and the deformed test (blue dashed line) signal. Ten iterations are illustrated. The bottom panel shows the growth of the displacement field after each iteration.

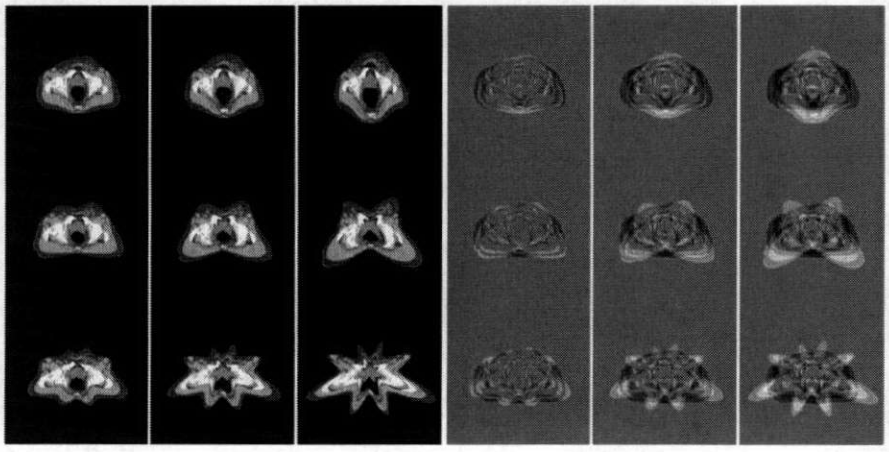


Figure 3. Synthetically deformed test images and their differences from the reference image. The left three columns show the test images. The right three columns show the difference images. All these test images are deformed from the same one slice CT (512 × 512) according to a harmonic deformation formula. From the top row to the bottom row, the harmonic parameter was $m = 2, 4, 8$ respectively. From left to right column, the harmonic parameter $b = 0.1, 0.2, 0.3$, respectively.

Here $\theta = \tan^{-1}\frac{y}{x}$. Two parameters m, b are used in the harmonic deformation, where m specifies the complexity of deformation, and $b$specifies the magnitude of deformation.

Figure 3 shows the deformed test images with different harmonic parameters. All these images are deformed from the same reference image. From top row to bottom row, the harmonic parameter $m = 2, 4, 8$ respectively. From left to right column, the harmonic parameter $b = 0.1, 0.2, 0.3$, respectively.

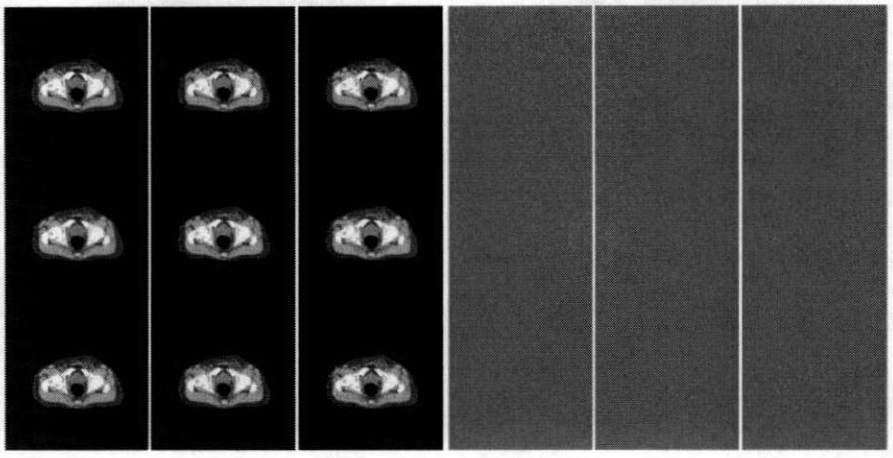


Figure 4. The registered version of the synthetic test images as illustrated in figure 3. The right three columns show the difference images after registration. The registrations use the technique presented in this paper.

We registered the test images to the reference image using the algorithm developed in this paper. A multi-resolution technique as discussed in section 3.2 is used. The Laplacian weight was set to be$\lambda = 0.1$, and 16 iterations were used for the finest level. For the comparison purpose, we also registered the same date using Thirion's 'demons' method (Thirion 1995, 1998, Ibanez *et al* 2003). Same multi-resolution strategies were used. As required by 'demons' method for the regularization purpose, a standard deviation ($\sigma = 1.0$) of the Gaussian smoothing kernel was applied to the demons' deformation field after each iteration. For registering 2D images of size 512×512 using proposed multi-resolution technique, the registration time is approximately 4 s when using our method and about 7 s when using 'demons' method, both were run on the same personal computer with clock speed of 933 MHz.

Figure 4 illustrates the registered version of figure 3 using our method. Figure 5 illustrates the registered version of figure 3 using 'demons' method. Figure 6 shows the similarity measures (CC) between the reference image and the test images before and after registration. The left panel is the CC before registration. The middle is the CC after registration using our method. The right panel shows the CC after registration using 'demons' method. For small and simple deformation ($m = 2, b \leqslant 1$), both our method and 'demons' method do a good job, the deformed images have been registered back to their reference counterpart. But for the large deformations ($m \geqslant 4, b > 1$), our method is superior to the 'demons' method. There is large registration error in 'demons' method. While our method does a great job, the similarity measures after registration approaches unity even for large and complex deformations ($m = 8, b = 0.3$).

5.2. Phantom data

To set-up a gold standard for testing the deformable registration algorithm, we designed a gel–balloon phantom (Lu 2001). In this phantom, a balloon was surrounded by the gel, and can be inflated and deflated with insertion or removal of heavy oil. The volume of balloon is controlled by the amount of oil injected, which can be read from an attached syringe. In total 320 plastic beads were implanted around the balloon in a regular cube grid ($7 \times 7 \times 7$,

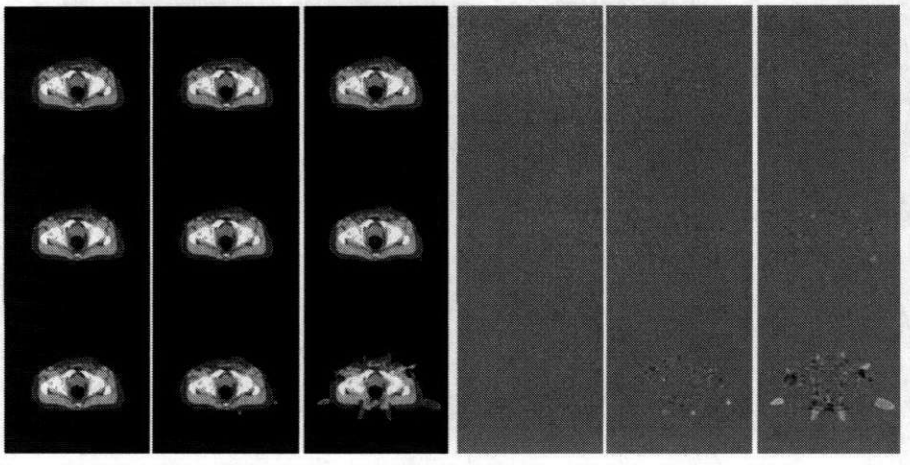


Figure 5. Similar to figure 4 but using 'demons' method for registration.

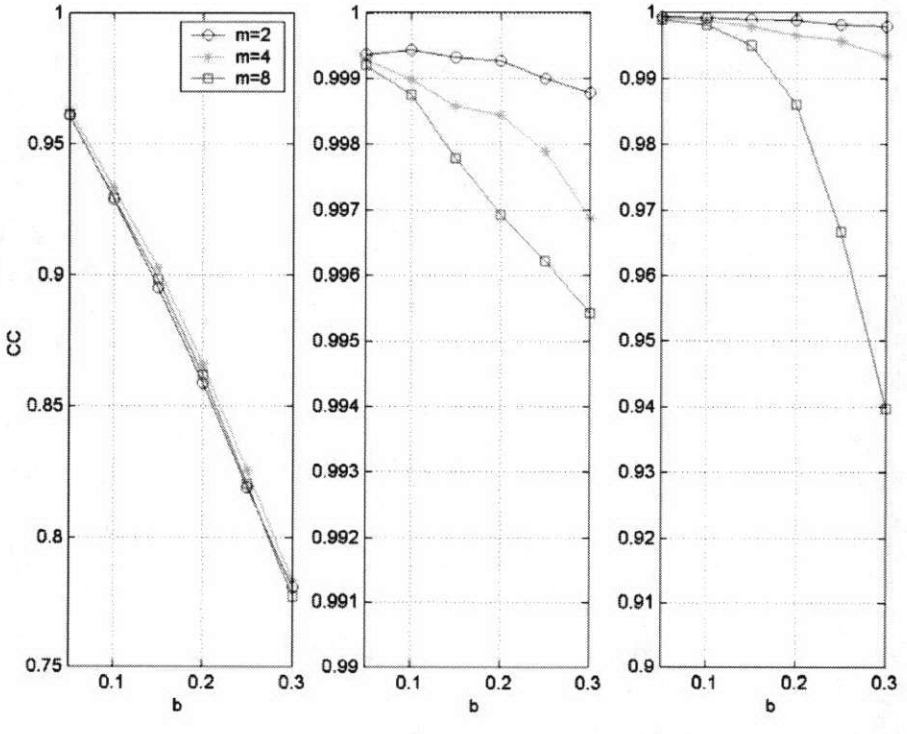


Figure 6. Similarity measures (CC) between reference image and test images. The left panel shows CC before registration. The middle panel shows CC after registrations based on the technique presented in this paper. The right panel shows CC after registrations based on the 'demons' method.

except the locations taken up by the balloon). The diameter of each bead is about 3.5 mm. The distances between beads were about 10 mm in each direction. The beads were assumed to move in the same way as the surrounding gel since they are moulded into the gel. By adjusting the amount of oil injected into the balloon, five deformation stages were approached. We use the medial deformation stage as the reference (R) and the other four stages are categorized as either inflation (I) or the deflation (D) stages. Thus, as in the increasing order of balloon size, the five deformation stages are D2, D1, R, I1 and I2 (1 and 2 stands for two levels of deformation).

CT images of the gel–balloon phantom were taken in order to study the movement and deformation of the gel as a function of the inflation state of the balloon. A Somatom HiQ CT scanner (Siemens Medical Systems, Inc., New Jersey, USA) at the University of Wisconsin Hospital and Clinics was used. The scan protocol was 140 KVp, 95 mAs, with voxel size $0.85 \times 0.85 \times 2\ \text{mm}^3$.

Based on the CT image, the positions of the beads can be measured, either manually or automatically. They can also be calculated by the deformable registration algorithm. The evaluation of the algorithm includes comparing these two results: measurement and calculation. Since the deformation is quite small (of the order of 1 mm) in this experiment, the measurement must be in sub-voxel resolution. We use an automatic algorithm (Lu 2001) to determine the positions of the beads.

Figure 7 shows the overlapped orthogonal (transverse, sagittal and coronal) view of the reference stage and that of inflated stage. The bottom (red) is the reference stage, while the top (green) is the inflated stage. The inflation of the balloon pushed the beads out, which is demonstrated in all transversal, coronal and sagittal views. Figure 8 illustrates the registered version of figure 7. It shows that both balloon and beads are registered to their reference position, and no significant difference between the reference and inflated stages is visible. Figures 9 and 10 are similar to figures 7 and 8, except now a deflation stage is registered with the reference stage.

The cumulative histograms of beads displacement are illustrated in figure 11. The y-axis of histogram shows the percentage of beads that has displacement smaller than certain value. The x-axis shows the amount of displacement. The left panel shows the histogram before registration, the right panel shows the histogram after registration. From the histogram plots, we found that the displacement of beads between the inflation (deflation) stages and the reference stage could be 3–4 mm. But after the registration, all displacements are less than 1 mm. This means that for small deformation, the registration error is less than 1 mm.

5.3. Clinical data

'Fractionation' (ICRU 1993, 1999) is a common technique in external beam radiotherapy; that is, a prescribed dose is delivered through multiple (such as 30–40) fractions, usually one fraction per day. The delivery time for each fraction lasts several minutes. We test our algorithm on both 'intra-fraction' images and 'inter-fraction' images. 'Intra-fraction' images stand for images that reveal 'intra-fraction' deformation, examples are CT images of different breathing phases. 'Inter-fraction' images stand for images that reveal daily deformation, such as daily CT images of the same patient during whole radiotherapy treatment courses.

5.3.1. Intra-fraction image registration. In this experiment, we use the CT data from a lung patient with IRB approval and de-identification. The CT scans were taken at University of Wisconsin-Madison Hospital and Clinics using a GE LightSpeed CT scanner with four row detectors (GE Medical System, Waukesha, WI). CT images at the exhalation phase and deep

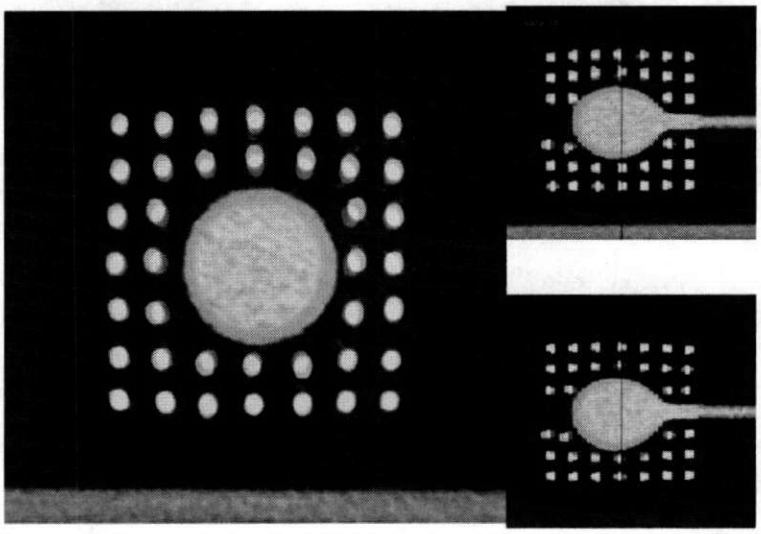


Figure 7. Overlapped orthogonal views of the reference stage (red) and one of the inflated stage (green) before registration. The left panel shows the transverse view, and the right panels show coronal (lower) and sagittal (upper) view, respectively.

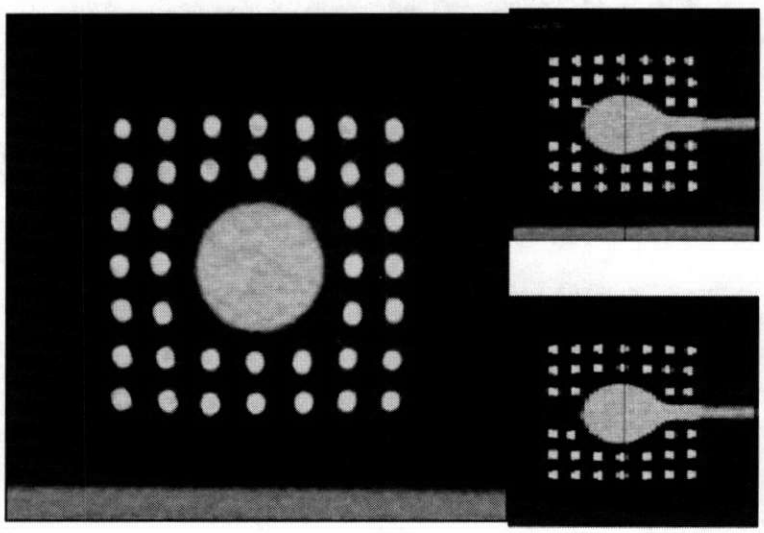


Figure 8. The registered version of figure 7. The red component is the reference stage and the green component is the registered test stage.

inspiration phase were obtained by breath-holding technique. The scan at deep inspiration phase took about 20 s with 0.8 s per gantry rotation. The exhalation phase was scanned with 0.4 s per gantry rotation and finished within 10 s. Both scans are helical with pitch = 1.5. The voxel size is $0.9 \times 0.9 \times 2.5$ mm^3.

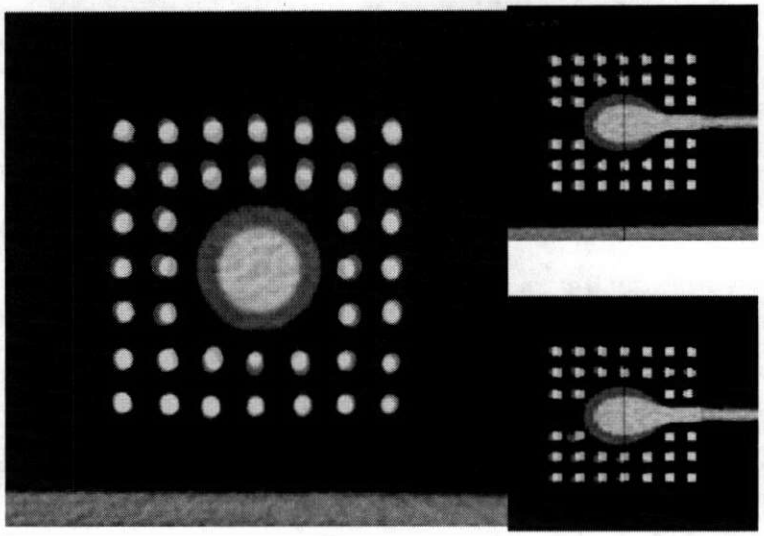


Figure 9. Overlapped orthogonal views of the reference stages (red component) and one of the deflated stages (green component) before registration. The left panel shows the transverse view, and the right panels show coronal (lower) and sagittal (upper) view, respectively.

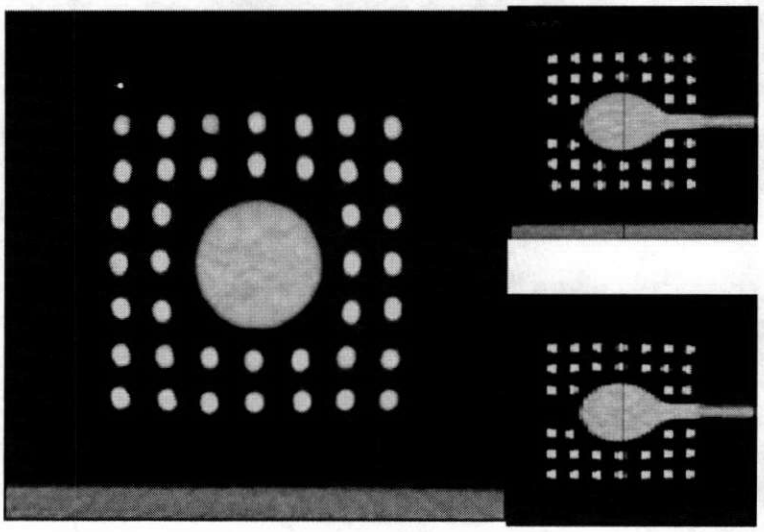


Figure 10. The registered version of figure 9. Red component is the reference stage and green component is the registered test stage.

CT images of two respiration stages, namely, inspiration and deep exhalation are used for the registration test. The inhale image is used as the reference image. The exhale stage is used as the test image. Figure 12 shows the overlapped orthogonal view of the inhale and exhale of CT image before deformable registration. The red component is the inhale image and the green component is the exhale image. We can see that there is large lung deformation from

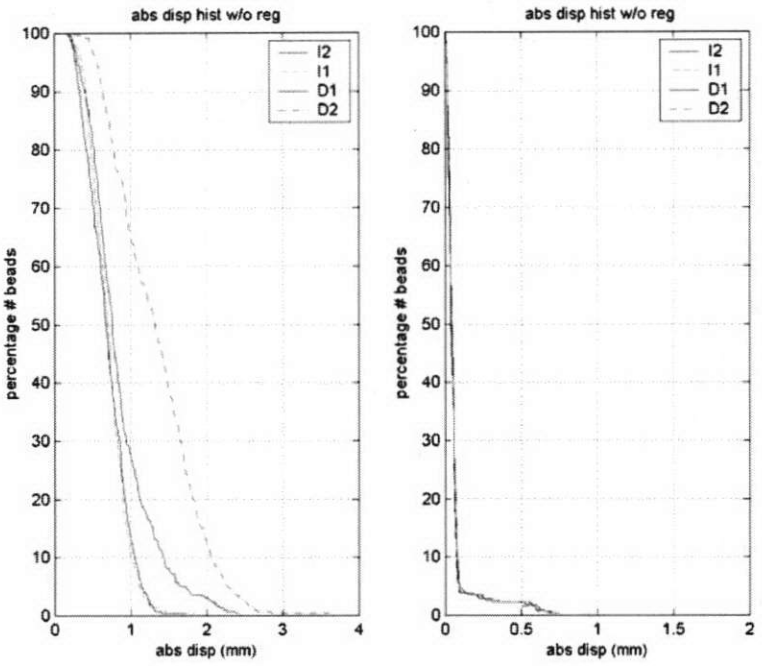


Figure 11. Histograms of absolute beads displacement before (left) and after (right) the registration.

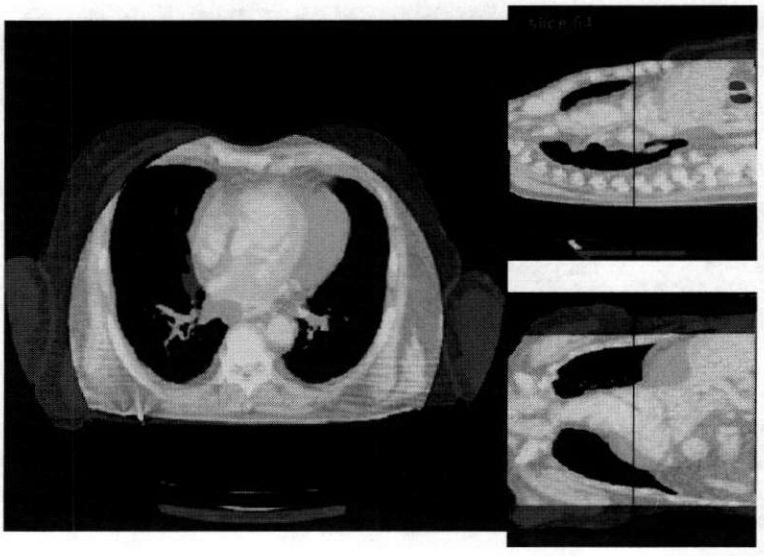


Figure 12. Overlapped orthogonal views of the intra-fraction CT images. The red component is the inhale image, which is used as reference image. The green component is the exhale image, which is used as test image.

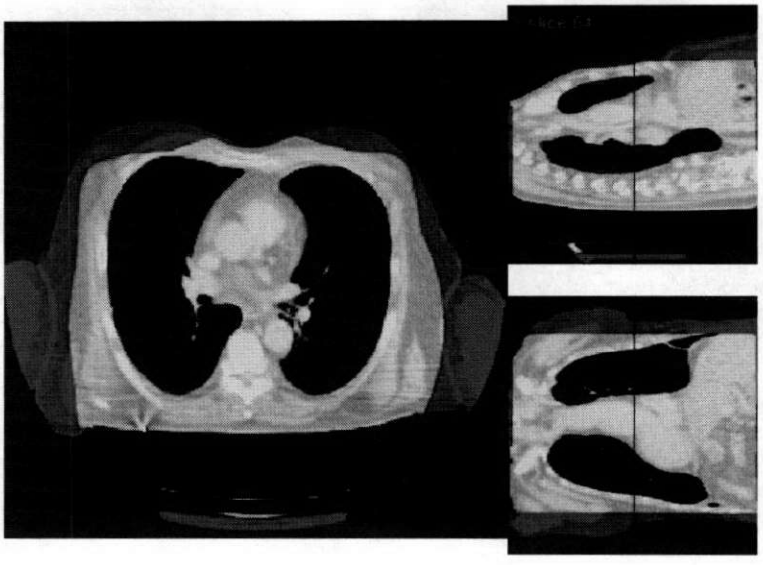


Figure 13. Overlap of inhale and the exhale image after registration.

inhale to exhale stages. In addition, the exhale data are reconstructed using an incomplete field of view. Figure 13 illustrates the overlapped orthogonal CT views of the inhale and exhale phases after deformable registration. The large deformation of lung and other parts of the body is registered. Significant misregistration only occurs at the edge portion of image, which is due to an incomplete test data set. This experiment shows that our method is robust for large deformations.

5.3.2. Inter-fraction image registration. In this experiment, we retrospectively study the inter-fractionCT data from three prostate patients. These data were taken on alternate days of external beam radiation therapy course in William Beaumont Hosiptal (courtesy Dr Di Yan) in 1997. The numbers of CT data set for these patients are 18, 15 and 16, respectively. For each patient, the first CT image set is treatment planning CT and all others are images on treatment days (fraction CT). We want to register the fraction images with the planning image. The planning CT is used as the reference image and the fraction CTs are used as the target image. All images are down-sampled to size of $256 \times 256 \times 61$, with voxel resolution $1.9 \times 1.9 \times 3\ \text{mm}^3$. The whole registration time between two images is about 3 min when run on a single processor 933 MHz Pentium III personal computer.

Figure 14 illustrates the overlapped orthogonal views of the reference image (bottom layer, red) and one of the rigid-body registered test images (top layer, green) of the first patient. The rigid-body registration is based on bony structures. It is still obvious that there is large local deformation between the reference images and the test images, most notably in the prostate and rectum region. The rectum is full in the reference image, while it is gas-filled in the test image. The prostate was deformed as well. Figure 15 illustrates the deformable registered version of figure 14. It shows that the local deformation is largely recognized and registered. There is hardly any significant difference between the reference image and thereak registered image. Figure 16 shows the enlarged view of two transverse slices from figures 14 and 15. The top is before deformable registration. The bottom is after deformable registration.

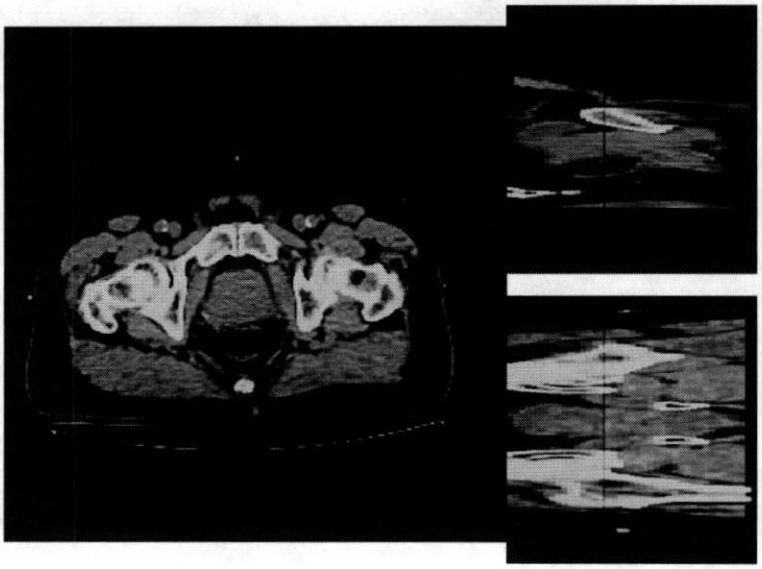


Figure 14. Overlapped orthogonal view of the CT images of the same prostate patient acquired during different days. The red component is the planning CT image, which is used as the reference image. The green component is the CT image acquired during the treatment day, which is used as the test image.

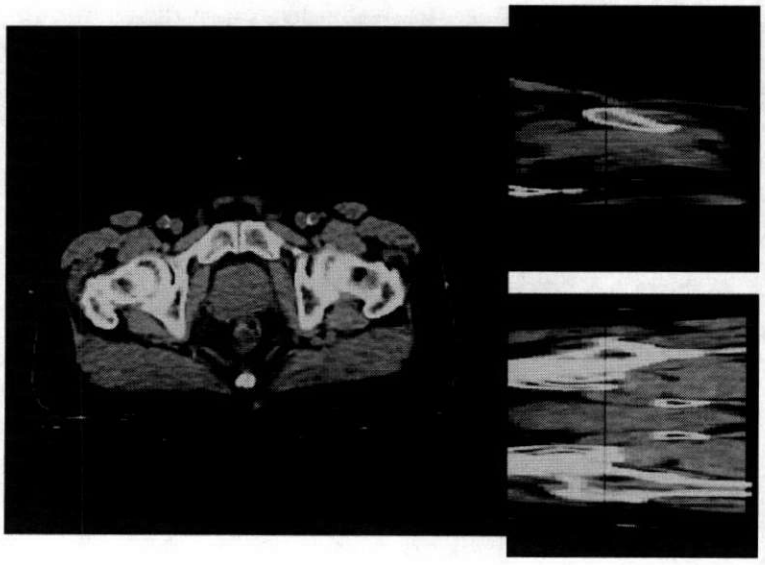


Figure 15. Overlap orthogonal views of planning and fraction image after registration.

Figure 17 compares similarity measures of cross correlation (CC) and mutual information (MI) after rigid-body registration and after deformable registration on three prostate patient

382

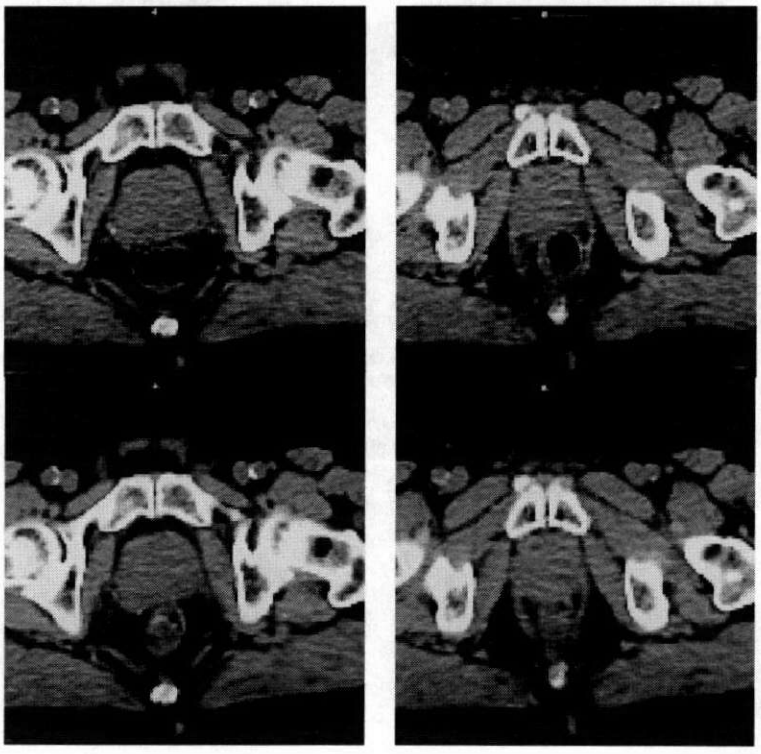


Figure 16. Enlarged view of two transverse slices as shown in figures 14 and 15. The top is before registration. The bottom is after registration.

data sets. The first row is CC and the second row is MI. The 'squares' represent similarity measures after the rigid-body registration but prior to the deformable registration. The 'circles' represent the similarity measures after the deformable registration. Prior to the deformable registration, the fraction images and the planning image have low similarity measures, which imply that there are large misalignments between these images. The similarity measures after deformable registration are significantly larger than that after rigid-body registration. It means that deformable registration is an indispensable step to account for local deformation.

6. Discussions and future work

As any voxel intensity similarity (sum of squared distance) based methods, the algorithm developed in this paper relies on the assumption that voxel representing the same homologous point on an object has the same intensity on both the reference and the test images. This assumption is reasonable for some applications in radiotherapy treatment planning and evaluation, where deformable registration of same modality images (daily CT data and 4D CT data) is required. This assumption prohibits this technique to be applied directly to multi-modality registration. But this limitation can somehow be bypassed by certain pre-processing steps. For example, the voxel intensity of both reference and test images can be mapped to physical density before the registration. The intensity mapping can also be achieved by some adaptive intensity correction technique (Guimond *et al* 2001). Based on these pre-processing

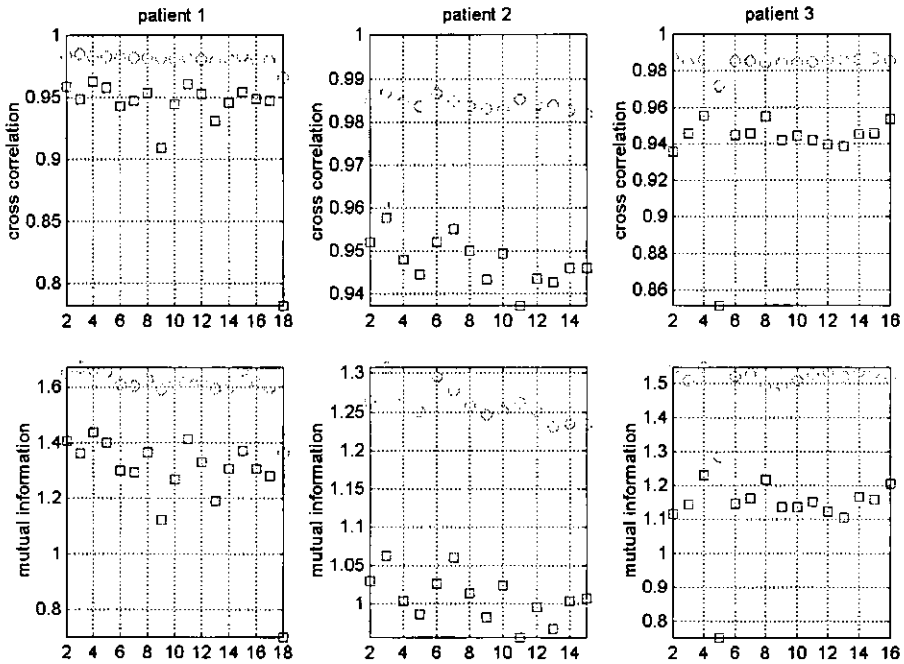


Figure 17. Similarity measures between planning CT (reference image) and fraction CT (test images) before (square) and after (circle) deformable registration for three patient images. The top row shows cross correlation, the bottom row shows mutual information. The numbers of CT set for each patient are 18, 15 and 16, respectively.

steps, this method could be potentially extended to multi-modality image registration, such as registration of KV-CT versus MV-CT, CT versus MRI, etc. On the other hand, using the mutual information (Likar and Pernus 2001) rather than the residual difference in the first part of equation (2.3) could be a more general solution. All these extensions need further investigations.

One problem common to any regularization-based method is how to choose the regularization parameter, the Laplacian weight λ as in our algorithm. Though through multi-resolution technique, the dependence on λ is much alleviated in our algorithm. A rather conserved choice of λ is enough for most applications as in radiation therapy planning and evaluation. Adaptively selecting λ for different iterations and/or tissues is worth a further investigation for the applications that require more accurate control of local deformation.

Another interesting topic of deformable registration is how to map points that are missing in the test image to that structure in the reference image. The driving force of our updating scheme (equations (3.10) and (3.1)) is $\lambda\nabla^2(u_{m,n}) - [B^*_m - A_m]g^*_{m,n}$; that is, if an object is missing in the test image, then $g^* \approx 0$ at that region, therefore, the displacement of such region is solely determined by the Laplacian term. The resulting displacement field of such region is slowly varying in accordance with its neighbour regions. This feature follows the standard principle of making the result smooth when there are no data. This topic will be further studied in the future work.

7. Conclusions

A fully automatic, fast and accurate deformable registration algorithm is presented. We tested the algorithm extensively using simulation data, phantom data, as well as the clinical data. Experimental results show that our model and algorithm are suitable for temporal deformable registration of same modality images. Simulation and phantom studies show that registration error of our method is sub-voxel. When applied to the intra-fraction CT images of a lung patient and inter-fraction image of a prostate patient. The simulation study and the lung patient study show that our method is suitable to account for large deformations. The prostate results show that registration based on local free-form deformation with smoothness constrains is appropriate to describe the local deformations in the prostate and the rectum. Similarity measures improved significantly after the registration.

Acknowledgments

The authors would like to thank Dr Di Yan of Beaumont Hospital to provide the prostate patient CT data, Tiezhi Zhang of University of Wisconsin-Madison to provide the lung patient CT data, Dr Harry Keller and Stewart Becker of University of Wisconsin-Madison to help on the phantom experiments.

References

Bajcsy R and Kovacic S 1989 Multiresolution elastic matching *Comput. Vis. Graph. Image Process.* **46** 1–21

Bookstein F 1989 Principal warps: Thin-plate spline and the decomposition of deformations *IEEE Trans. Pattern Anal. Mach. Intell.* **11** 567–85

Brock K, Balter J, Dawson L, Kessler M and Meyer C 2003 Automated generation of a four-dimensional model of the liver using warping and mutual information *Med. Phys.* **30** 1128–33

Brown L G 1992 A survey of Image registration techniques *ACM Comput. Surv.* **24** 325–76

Burr D J 1981 A dynamic model for image registration *Comput. Graph. Imaging Process.* **15** 102–12

Christensen G E, Rabbitt R D and Miller M I 1996 Deformable templates using large deformation kinematics *IEEE Trans. Image Process.* **5** 1435–47

Gee J C, Antoine Maintz J B and Vannier M W 2003 *Biomedical Image Registration: Revised Papers. Second International Workshop, WBIR 2003 (Lecture Notes in Computer Science)* (Philadelphia, PA: Springer)

Guimond A, Roche A, Ayache N and Meunier J 2001 Three-dimensional multimodal brain warping using the Demons algorithm and adaptive intensity corrections *IEEE Trans. Med. Imaging* **20** 58–69

Hagemann A, Rohr K, Stiehl H S, Speteger U and Gilsbach J M 1999 Biomechanical modeling of the human head for physically based, nonrigid image registration *IEEE Trans. Med. Imaging* **18** 875–84

Hajnal J V, Hawkes D L G and Hill D J (eds) 2001 *Medical Image Registration* (Boca Raton, FL: CRC Press)

Hellier P, Barillot C, Memin E and Perez P 2001 Hierarchical estimation of a dense deformation field for 3D robust registration *IEEE Trans. Med. Imaging* **20** 388–402

Hill D L G, Batchelor P G, Holden M and Hawkes D J 2001 Medical image registration *Phys. Med. Biol.* **46** 1–45

Horn B and Schunck B 1981 Determining optical flow *Artif. Intell.* **17** 185–203

Ibanez L, Schroeder W, Ng L and Cates J 2003 *The ITK Software Guide: The Insight Segmentation and Registration Toolkit* (New York: Kitware Inc.)

ICRU 1993 Prescribing, recording, and reporting photon beam therapy *ICRU Report 50*

ICRU 1999 Prescribing, recording, and reporting photon beam therapy *Supplement to ICRU Report 50*

Keener J P 1988 *Principles of Applied Mathematics: Transformation and Approximation* (Reading, MA: Addison-Wesley)

Kyriacou S K and Davatzikos C 1999 *A 3D model of quasi-static brain mechanics: application to tumor-bearing image registration 1st Joint BMES/EMBS Conference* p 1161

Kyriacou S K, Davatzikos C, Zinreich S J and Bryan R N 1999 Nonlinear elastic registration of brain images with tumor pathology using a biomechanical model [MRI] *IEEE Trans. Med. Imaging* **18** 580–92

Lester H and Arridge S R 1999 A survey of hierarchical nonlinear medical registration *Pattern Recognition* **32** 129–49

Likar B and Pernus F 2001 A hierarchical approach to elastic registration based on mutual information *Image Vis. Comput.* **19** 33–44

Lu W 2001 Motion detection and correction for image guided radiation therapy *PhD Thesis* University of Wisconsin-Madison

Lu W, Chen M and Mackie T R 2001 Deformable image registration: I. Theory *Biological & Physical Basis of IMRT & TomoTherapy* ed B R Paliwal, D E Herbert, J K Fowler and M P Mehta (Madison WI: Medical Physics Publishing) pp 322–31

Lu W, Keller H and Mackie T R 2001 Deformable Image Registration: II. Experiments *Biological & Physical Basis of IMRT & Tomotherapy* ed B R Paliwal, D E Herbert, J K Fowler and M P Mehta (Madison, WI: Medical Physics Publishing) pp 332–42

Lu W, Mackie T R, Keller H, Ruchala K J and Olivera G H 2000 *A generalization of adaptive radiotherapy and the registration of deformable dose distribution ICCR 2000* pp 521–3

Maes F, Dollignon A, Vandermeulen D, Marchal G and Suetens P 1997 Multimodality image registration by maximization of mutual information *IEEE Trans. Med. Imaging* **16** 187–98

Maintz J B A and Viergever M A 1998 A survey of medical imaging registration *Med. Image Anal.* **2** 1–36

Mattes D, Haynor D, Vesselle H, Lewellen T and Eubank W 2003 PET-CT image registration in the chest using free-form deformations *IEEE Trans. Med. Imaging* **22** 20–8

Miller M I, Christensen G E, Amit Y and Grenander U 1993 Mathematical textbook of deformable neuroanatomies *Proc. Natl. Acad. Sci. USA* **90** 11944–8

Olivera G H, Ruchala K, Lu W, Kapatoes J, Reckwerdt P, Jeraj R and Mackie R 2003 Evaluation of patient setup and plan optimization strategies based on deformable dose registration *Int. J. Radiat. Oncol. Biol. Phys.* **57** S188–9

Pluim J P W, Maintz J B A and Viergever M A 2003 Mutual-information-based registration of medical images: a survey *IEEE Trans. Med. Imaging* **22** 986–1004

Press W H, Teukolsky S A, Vetterling W T and Flannery B P 1992 *Numerical Recipes in C* (Cambridge: Cambridge University Press)

Rogelj P and Kovacic S 2001 Similarity measures for non-rigid registration *Proc. SPIE* **4322** 569–78

Sederberg T and Parry S 1986 Free form deformation of solid geometry models *SIGGRAPH* **20** 151–60

Studholme C, Hill D L G and Hawkes D J 1999 An overlap invariant entropy measure of 3D medical image alignment *Pattern Recognition* **32** 71–86

Thirion J P 1995 *Fast non-rigid matching of 3d medical images* (MRCAS'95: Medical Robotics and Computer Aided Surgery Baltimore) pp 47–54

Thirion J P 1998 Image matching as a diffusion process: an analogy with Maxwell's demons *Med. Image Anal.* **2** 243–60

van den Elsen P A, Pol E J D and Viergever M A 1993 Medical image matching-a review with classification *IEEE Eng. Med. Biol. Mag.* **1** 26–39

Wang Y and Staib L H 1998 *Integrated approaches to non-rigid registration in medical images IEEE WACV '98* pp 102–8

Woods R, Grafton S, Watson J, Sicotte N and Mazziotta J 1998 Automated image registration: II. Intersubject validation of linear and nonlinear models *J. Comput. Assist. Tomogr.* **22** 153–65

Xu C 2000 Deformable models with application to human cerebral cortex reconstruction in magnetic resonance images *PhD Thesis* John Hopkins University

Yan D, Jaffray D A and Wong J W 1999 A model to accumulate fractionated dose in a deforming organ *Int. J. Radiat. Oncol. Biol. Phys.* **44** 665–75

Zhang T, Jeraj R, Keller H, Lu W, Olivera G H, McNutt T R, Mackie T R and Paliwal B 2004 Treatment plan optimization incorporating respiratory motion *Med. Phys.* **31** 1576–86

Zitova B and Flusser J 2003 Image registration methods: a survey *Image Vis. Comput.* **21** 977–1000

Accurate Inverse Consistent Non-rigid Image Registration and Its Application on Automatic Re-contouring

Qingguo Zeng and Yunmei Chen

Department of Mathematics, University of Florida
Little Hall 358, Gainesville, FL 32611-8105
{qingguo, yun}@math.ufl.edu
http://www.math.ufl.edu/~{qingguo,yun}

Abstract. This paper provides a novel algorithm for invertible non-rigid image registration. The proposed model minimizes two energy functionals coupled by a natural inverse consistent constraint. Both of the energy functionals for forward and backward deformation fields consist a smoothness measure of the deformation field, and a similarity measure between the deformed image and the one to be matched. In this proposed model the similarity measure is based on maximum likelihood estimation of the residue image. To enhance algorithm efficiency, the Additive Operator Splitting (AOS) scheme is used in solving the minimization problem. The inverse consistent deformation field can be applied to automatic re-contouring to get an accurate delineation of Regions Of Interest(ROIs). The experimental results on synthetic images and 3D prostate data indicate the effectiveness of the proposed method in inverse consistency and automatic re-contouring.

1 Introduction

Image registration, a very important subject in computer vision and image processing, has been increasingly used in image guided surgery, functional brain mapping, multi-modality fusion etc. The task of image registration is to find a transformation h that relates points in the source image S to their corresponding points in the target image T. This transformation can be either rigid or non-rigid (deformable). Rigid registration is restricted to be a combination of scaling, rotation and translation only, hence, it is not adequate for applications involving large free deformations, for example, image guided radiation therapy on prostate. Deformable image registration allows more freedom at each point, it has been the subject of extensive study in the literature (e.g. [1,20,19,24,14,17,23,10,5]). In most deformable registration models, a smooth and "natural" deformation field h is usually driven by intensity based similarity measures such as Sum of Square Distance(SSD) ([14,16]), Cross Correlation (CC)([6,4]), Mutual Information (MI) ([22,21,6,18]) between the deformed source image $S(h(\mathbf{x}))$ and the target image $T(\mathbf{x})$.

I. Măndoiu, R. Sunderraman, and A. Zelikovsky (Eds.): ISBRA 2008, LNBI 4983, pp. 293–304, 2008.

In certain applications such as imaging guided radiation therapy, it would be better to have a one-to-one and inverse consistent deformation field, while the majority of non-rigid registration methods do not guarantee such property. The inverse consistent means that when the source and target images are switched in the model, the point correspondence between S and T does not change. An inverse inconsistent deformation field can generate large errors in the processes like auto re-contouring([16]), dose calculate ([11,15]) in radiation therapy. A number of work have attempted to make the registration inverse consistent (e.g. [3,12,2]). Here we only discuss two of them which are closely related to our work. In [3], Christensen and Johnson proposed the following coupled minimization problems:

$$\begin{aligned} E(h) &= \underbrace{M(S(h),T) + \lambda R(h)}_{E_1} + \rho \int_\Omega |h - g^{-1}|^2 dx \\ E(g) &= \underbrace{M(S,T(g)) + \lambda R(g)}_{E_2} + \rho \int_\Omega |g - h^{-1}|^2 dx \end{aligned} \tag{1}$$

where $M(\cdot,\cdot)$ is a dissimilarity measure between two images, g is the backward mapping which deforms T such that $T(g)$ is close to S under measure M, and g is expected to be the inverse of the forward mapping h (i.e. $h \circ g = g \circ h = id$, where id is the identity mapping). $R(\cdot)$ is a regularity measure on deformation fields h and g, $\lambda > 0$ and $\rho > 0$ are parameters balances the goodness of alignment, the smoothness of the deformation, and consistence of invertibility. g^{-1} and h^{-1} represent numerical inverses of g and h respectively. In [3], h and g were solved by using gradient descent algorithm.

Since the inverse consistent constraints are accommodated by penalty terms in the energy functionals, solutions h and g by (1) are not exactly inverse to each other. How h and g are closely inverse to each other depends on how large the parameter ρ is, which in practice is hard to choose and needs to be adjusted case by case. Theoretically, h and g are exactly inverse to each other only when $\rho \to +\infty$.

In [12], Leow et al. proposed a different approach. They find h and g by the time marching scheme:

$$h^{(n+1)} = h^{(n)} + dt(\eta_1 + \eta_2), \quad g^{(n+1)} = g^{(n)} + dt(\xi_1 + \xi_2), \tag{2}$$

where dt is the time step, η_1 and ξ_2 are vector fields representing gradient descent directions of E_1 and E_2 in (1) respectively, i.e.

$$\begin{aligned} \eta_1(x) &= -\nabla_h M(S(h(x)),T(x)) - \lambda \nabla_h R(h(x)) \\ \xi_2(x) &= -\nabla_g M(S(x),T(g(x))) - \lambda \nabla_g R(g(x)). \end{aligned} \tag{3}$$

To make the model inverse consistent, η_1 and ξ_2 are chosen by the following approach.

Suppose $h^{(n)} \circ g^{(n)} = id$ in the nth iteration, then η_2, ξ_1 were determined by taking care of the inverse consistent constraints $h^{(n+1)} \circ g^{(n+1)} = id$, i.e.

$$(h^{(n)} + dt(\eta_1 + \eta_2)) \circ (g^{(n)} + dt(\xi_1 + \xi_2)) = id. \tag{4}$$

Taking the Taylor's expansion of (4) with respect to dt at 0, and collecting up to the first order terms of dt, one gets

$$\eta_2(x) = -D(h(x))\xi_2(h(x)),\ \xi_1(x) = -D(g(x))\eta_1(g(x)) \tag{5}$$

where D is the Jacobian matrix operator. Relations in (5) make the iterations in (2) uni-directional, i.e., updating formula for the forward mapping h does not depend on the backward mapping g, and vise versa.

In this scheme, the driving force for updating h (or g) involves both forward force from E_1 and backward force from E_2, so the scheme aligns two images faster than the models in which the force driving the deformation field depends on E_1 or E_2 only([23]).

However, $h^{(n+1)}$ and $g^{(n+1)}$ by (2-5) are not exactly inverse to each other even $h^{(n)}$ and $g^{(n)}$ are. Since in the derivation of (5), the higher order terms in the Taylor expansion of the left hand side of (4) have been discarded. This generates truncation errors, which are accumulated and exaggerated during iterations. Started with the identity mapping for both h and g, the solutions h and g from (2-5) are not inverse to each other, as we will show in experimental results.

Regarding the dissimilarity measure $M(\cdot,\cdot)$, a conventionally used one for same modality image registration is SSD, which is sensitive to the presence of noise and outliers (e.g. [7,8]). Moreover, the fixed parameter λ in (1) balancing the smoothness of the deformation field and goodness of the alignment is always difficult to select, and affects the robustness of the model to the choice to this weighting parameter. Small λ results an unstable and discontinuous deformation field, while large λ leads to inaccurate result, and may yield a nonphysical deformation field due to unreasonable restrictions. In our proposed model we will replace the SSD dissimilarity measure by a likelihood estimation that is based on the assumption of a Gaussian distribution of the residue image.

The main contribution of this work is on the improvement of inverse consistency, and its application to the radiation therapy, in particular, to get more accurate auto re-contouring. Our basic idea is minimizing E_1 and E_2 in (1) coupled by the inverse consistent constraints defined in the next section. Applications of these inverse consistent deformations on auto re-contouring will be discussed in experimental results.

2 Proposed Method

In this section, we will first introduce a "natural" formulation of inverse consistent constraints, which can be used to correct the truncation errors in (4). Then we will combine this with the dissimilarity measures based on the likelihood of the residue image into our energy functionals to improve the accuracy, robustness and inverse consistent of the deformable image registration.

Let u and v be the forward and backward displacement fields related to deformation h and g by

$$h(\mathbf{x}) = \mathbf{x} + \mathbf{u}(\mathbf{x}),\ \ g(\mathbf{x}) = \mathbf{x} + \mathbf{v}(\mathbf{x}). \tag{6}$$

Then the inverse consistent constraint $h(g(\mathbf{x})) = \mathbf{x}$ can be written in terms of u and v as

$$\mathbf{x} = h(g(\mathbf{x})) = g(\mathbf{x}) + u(g(\mathbf{x})) = \mathbf{x} + v(\mathbf{x}) + u(\mathbf{x} + v(\mathbf{x})). \tag{7}$$

Therefore,

$$v(\mathbf{x}) = -u(\mathbf{x} + v(\mathbf{x})) \tag{8}$$

Similarly, the constraint $g(h(\mathbf{x})) = \mathbf{x}$ can be represented by

$$u(\mathbf{x}) = -v(\mathbf{x} + u(\mathbf{x})) \tag{9}$$

In this work, we will use (8) and (9) as hard constraints in our proposed energy minimization method.

To accommodate certain degree of variability in the image matching, we consider the residue between the deformed source image and target image $S(h(\mathbf{x})) - T(\mathbf{x})$ at each point as an independent random variable with Gaussian distribution of mean zero and a variance σ to be optimized. By the independency assumption, the joint pdf of all these random variables, which is the likelihood of the residual image given parameter σ, becomes

$$\begin{aligned} p(\{S(h(\mathbf{x})) - T(\mathbf{x}), \mathbf{x} \in \Omega\}|T(\mathbf{x}), \sigma) &= \Pi_{\mathbf{x}\in\Omega} p(S(h(\mathbf{x})) - T(\mathbf{x})|T(\mathbf{x}), \sigma) \\ &= \Pi_{\mathbf{x}\in\Omega} \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{|S(h(\mathbf{x})) - T(\mathbf{x})|^2}{2\sigma^2}}. \end{aligned} \tag{10}$$

Then the negative log-likelihood function is

$$\int_\Omega \frac{|S(h(\mathbf{x})) - T(\mathbf{x})|^2}{2\sigma^2} d\mathbf{x} + |\Omega| \ln \sqrt{2\pi}\sigma \tag{11}$$

Replace $M(S(h), T)$ in E_1 of (1) by (11), and replace $M(S, T(g))$ in E_2 in a similar manner, E_1 and E_2 can be rewritten in terms of u and v as:

$$\begin{aligned} E_1(u) &= \int_\Omega \frac{|S(\mathbf{x} + u(\mathbf{x})) - T(\mathbf{x})|^2}{2{\sigma_1}^2} d\mathbf{x} + |\Omega| \ln \sigma_1 + \lambda R(u(\mathbf{x})) \\ E_2(v) &= \int_\Omega \frac{|S(\mathbf{x}) - T(\mathbf{x} + v(\mathbf{x}))|^2}{2{\sigma_2}^2} d\mathbf{x} + |\Omega| \ln \sigma_2 + \lambda R(v(\mathbf{x})) \end{aligned} \tag{12}$$

After choosing $R(\bullet) = |\bigtriangledown \bullet|^2_{L^2(\Omega)}$ with boundary and initial conditions:

$$\begin{aligned} &\frac{\partial u}{\partial \boldsymbol{n}}(x,t) = 0, \ \frac{\partial v}{\partial \boldsymbol{n}}(x,t) = 0, \ \text{on } \partial\Omega \times R^+ \\ &u(x,0) = 0, \ v(x,0) = 0, \ \text{on } \Omega, \end{aligned} \tag{13}$$

ξ_1 and η_2 in (3) become

$$\begin{aligned} \eta_1(x) &= \frac{T(x) - S(x + u(x))}{\sigma_1^2} \nabla S(x + u(x)) + \lambda \triangle u(x) \\ \xi_2(x) &= \frac{S(x) - T(x + v(x))}{\sigma_2^2} \nabla T(x + v(x)) + \lambda \triangle v(x) \end{aligned} \tag{14}$$

Taking the first variation of the energy functional (12), we get

$$\sigma_1^2 = \frac{\int_\Omega |S(x+u(x)) - T(x)|^2 dx}{|\Omega|},\ \sigma_2^2 = \frac{\int_\Omega |S(x) - T(x+v(x))|^2 dx}{|\Omega|}. \quad (15)$$

By using this dissimilarity measure, the residual image no longer needs to be pointwisely close to zero to make the L^2 norm small. Instead, the new measure only forces the mean of the residue to be zero, and allows the residue having a variance to accommodate certain variability. This is especially good for aligning two images whose intensities are not exactly equal or linearly related, and makes the model more robust to noise and artifacts.

Moreover, the likelihood based approach is less sensitive to the choice of the parameter λ. In SSD models, λ is prefixed, so the balance of the dissimilarity measure and regularity measure does not change during iterations. This makes the selection of λ very difficult and affects registration result. In the proposed model the balancing factor of these two measures is, in fact, $\lambda\sigma^2$ rather than λ alone. Therefore, even λ is prefixed, the weight between these two measures varies at each iteration as the variance updates. As the iterations gradually approach to convergence stage, the residue magnitude becomes smaller, hence, the variance σ reduces, and consequently, the weight on smoothing deformation field versus matching images automatically decreases.

Combining these ideas, we propose a new model to improve the inverse consistency of (1) by using the hard constraints (8) and (9) to replace the penalty terms in (1), and to improve the efficiency of alignment by using the proposed similarity measure. More precisely, we propose to minimize a coupled minimization problem

$$\begin{cases} min_{u\in W^{1,2}(\Omega),\sigma_1}\ E_1(u,\sigma_1) \quad \text{and} \quad min_{v\in W^{1,2}(\Omega),\sigma_2} E_2(v,\sigma_2) \\ \qquad\qquad \text{where } E_1, E_2 \text{ are defined in (12)} \\ \qquad\qquad subject\ to\ u(x) + v(x+u(x)) = 0\ \forall x \in \Omega \\ \qquad\qquad\quad v(x) + u(x+v(x)) = 0\ \forall x \in \Omega \end{cases} \quad (16)$$

Note if we only minimize $E_1(u)$, the solution u may not be invertible, thus we may not be able to get its inverse through (8). By choosing η_1 and ξ_2 as in (14), the solutions of the constrained coupled minimization problem (16) are obtained via the following algorithm

Algorithm 1

$u(x) = 0; v(x) = 0; iter = 0;$
$corr = correlation(S,T)$
while $iter < maxstep$ and $corr < threshhold$ **do**
  $iter++;$
  $\mathbf{u_{new}} = \mathbf{u} + \mathbf{dt}\eta_1(\mathbf{u})$
  $\mathbf{v} \leftarrow -\mathbf{u_{new}}(\mathbf{x} + \mathbf{v})$
  $\mathbf{v_{new}} = \mathbf{v} + \mathbf{dt}\xi_2(\mathbf{v})$
  $\mathbf{u} \leftarrow -\mathbf{v_{new}}(\mathbf{x} + \mathbf{u_{new}})$

$\mathbf{v} = \mathbf{v_{new}}$
$corr1 = correlation(S(x + u(x)), T(x))$
$corr2 = correlation(S(x), T(x + v(x)))$
$corr = min(corr1, corr2)$
end while

By applying the flow equations alternatively with the exact inverse consistent constraints in Algorithm 1, the forward and backward deformation forces are related via the constraints. In this manner, images to be registered are aligned well, meanwhile, the inverse inconsistency errors are controlled by hard constraints. The performance on image matching and improvement of accuracy on the inverse consistency will be shown in the next section.

To enhance the efficiency of the proposed algorithm, the Additive Operator Splitting scheme([13,9]) is used to speed up our numerical computation. We present our results based on synthetic images, and 3D prostate MRI data.

3 Experiment Results

In this section, we show our experimental results on 2D synthetic images and 3D prostate MRI data, which indicate the improvement of the proposed algorithm in accuracy of inverse consistency and accuracy of auto re-contouring.

Based on (8) and (9), if h and g are inverse to each other, then both of the inverse inconsistent error fields $u(\mathbf{x}) + v(\mathbf{x} + u(\mathbf{x}))$ and $v(\mathbf{x}) + u(\mathbf{x} + v(\mathbf{x}))$ should be 0. Thus, their components and norms should also be 0. This property will be used to compare the inverse inconsistency errors between three methods: Scheme (2), Algorithm 1 and Model (16) without constraints.

The first experiment is aimed to exam and compare the inverse consistency of these three methods on synthetic data. Fig 1(a) and (b) present the source image S and target image T, respectively, with the boundaries of the objects superimposed. Objects have intensity 1.0 in both images, and their background/holes have intensity 0. The three methods are applied separately with the same parameters $dt = .05, \lambda = 5.0$ for 800 iterations, and the corresponding results are shown in Fig 1(c-g). From Fig 1(c), one can see that correlation by all methods converges to about 1.0 similarly. However in Fig 1(d), the means of norms for both inverse inconsistent error fields $u(\mathbf{x}) + v(\mathbf{x} + u(\mathbf{x}))$ and $v(\mathbf{x}) + u(\mathbf{x} + v(\mathbf{x}))$ by both Scheme (2) and non-constrained Model (16) are increased to about 0.4 pixels in average, i.e., the mean value of their inverse inconsistency errors are about 0.4 pixels, while those from Algorithms 1 are maintained in a negligible low level (about 0.01 pixel). We also compare the error field $u(\mathbf{x}) + v(\mathbf{x} + u(\mathbf{x}))$ in Fig 1(e,f,g) by applying results from Scheme (2), Algorithm 1 and Model (16) without constraints respectively on a regular grid mesh. Fig 1(f) shows that the error by Algorithm 1 is 0 almost everywhere(almost no displacement in the regular grid mesh), while (e) and (g) show errors from other two methods are larger, especially in the region corresponding to the boundaries of two holes in S and T.

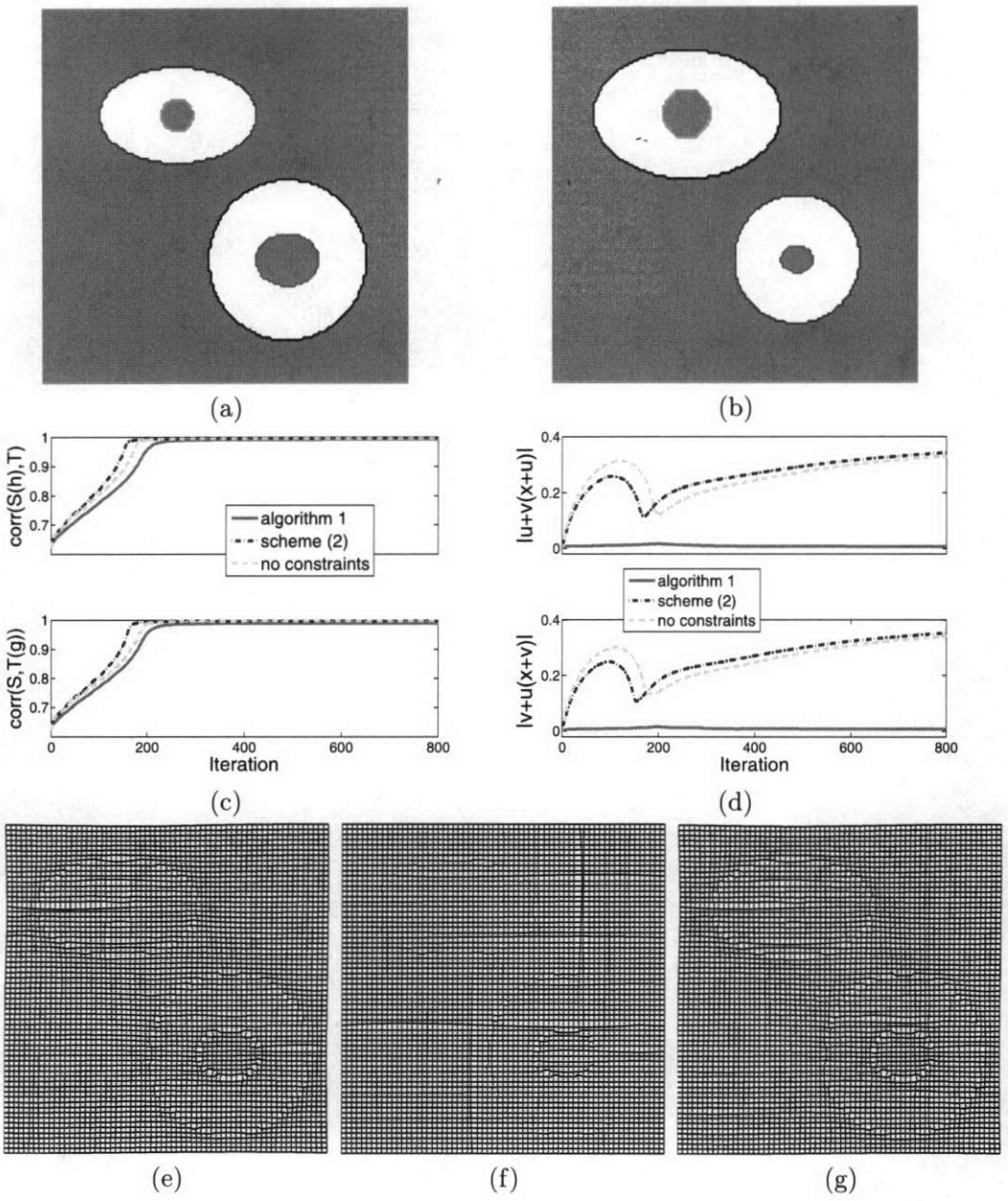


Fig. 1. Experiment on 2D synthetic images. (a) Source with object contours superimposed, (b) target with contours superimposed, (c)correlation $Corr(S(h), T)$ and $Corr(S, T(g))$ during iterations, (d) mean of norms $\|u(x) + v(x + u(x))\|_{L^2(\Omega)}$ and $\|v(x) + u(x + v(x))\|_{L^2(\Omega)}$ during iterations, (e-g) grid representations of inverse inconsistent error field $u(x) + v(x + u(x))$ by Scheme (2), Algorithm 1 with and without constraints respectively.

To quantitatively validate the improved inverse consistency by the proposed algorithm, we compare the maximum and mean values of components and norms of $u(\mathbf{x}) + v(\mathbf{x} + u(\mathbf{x}))$ and $v(\mathbf{x}) + u(\mathbf{x} + v(\mathbf{x}))$ in Table 1, where X_{max}, X_{mean} denote the maximum and mean values of the first component of the error fields

respectively, and Y_{max}, Y_{mean} are those of the second components. The quantitative comparisons are performed in two regions: one is in the image domain Ω, and the other is on the contours which are the boundaries of the objects in the images shown in Fig 1(a,b) respectively. Table 1 shows the proposed algorithm yields much smaller errors in all aspects. Particularly, its mean error is about one fortieth of those from both scheme (2) and non-constrained model (16) in Ω, and about one thirtieth of theirs at the contour regions.

The second experiment is to validate the improvement in accuracy of the proposed algorithm on 3D prostate data, which consists of 100 phases of 2D images

Table 1. Inverse inconsistency error comparison results for synthetic images: the components and norms of inverse inconsistency error fields $u(\mathbf{x}) + v(\mathbf{x} + u(\mathbf{x}))$ and $v(\mathbf{x}) + u(\mathbf{x} + v(\mathbf{x}))$. X_{max}, X_{mean} denote the maximum and mean values of the first component of the error fields respectively, Y_{max}, Y_{mean} denote that for the second component. $\| \bullet \|$ denotes norms of the error fields at each pixel. Ω is the image domain.

inconsistency error $u(\mathbf{x}) + v(\mathbf{x} + u(\mathbf{x}))$ on Ω						
Method	X_{max}	X_{mean}	Y_{max}	Y_{mean}	$\| \bullet \|_{max}$	$\| \bullet \|_{mean}$
Algorithm 1	0.7396	0.0051	0.7876	0.0059	0.7879	0.0085
Scheme (2)	0.9915	0.2390	1.1055	0.2320	1.2272	0.3526
No constraints	0.9916	0.2300	1.0979	0.2255	1.2202	0.3414
inconsistency error $v(\mathbf{x}) + u(\mathbf{x} + v(\mathbf{x}))$ on Ω						
Algorithm 1	0.4870	0.0047	0.5814	0.0053	0.7287	0.0076
Scheme (2)	0.8705	0.2290	1.2224	0.2240	1.2304	0.3398
No Constraints	0.8303	0.2199	1.2670	0.2174	1.2729	0.3284
inconsistency error $u(\mathbf{x}) + v(\mathbf{x} + u(\mathbf{x}))$ on contours in S						
Algorithm 1	0.2810	0.0431	0.2769	0.0502	0.3742	0.0719
Scheme (2)	0.7237	0.1445	0.8047	0.1878	0.8073	0.2577
No Constraints	0.7403	0.1461	0.8374	0.1908	0.8410	0.2612
inconsistency error $v(\mathbf{x}) + u(\mathbf{x} + v(\mathbf{x}))$ on contours in T						
Algorithm 1	0.3833	0.0412	0.4058	0.0484	0.4059	0.0697
Scheme (2)	0.5937	0.1390	0.9308	0.1661	0.9715	0.2361
No Constraints	0.6180	0.1410	0.9735	0.1695	1.0171	0.2402

Table 2. Inverse inconsistency error comparisons for the $1st$ and $21st$ phases of a 3D prostate MRI data on regions Ω and all the contours in the $1st$ phase S

Method	X_{max}	X_{mean}	Y_{max}	Y_{mean}	$\| \bullet \|_{max}$	$\| \bullet \|_{mean}$
inconsistency error field $u(\mathbf{x}) + v(\mathbf{x} + u(\mathbf{x}))$ on Ω						
Algorithm 1	0.5155	0.0105	0.4010	0.0082	0.5549	0.0146
Scheme (2)	0.4936	0.0457	0.6750	0.0793	0.7154	0.1020
inconsistency error field $v(\mathbf{x}) + u(\mathbf{x} + v(\mathbf{x}))$ on Ω						
Algorithm 1	0.5937	0.0121	0.4766	0.0087	0.6230	0.0164
Scheme (2)	1.6315	0.0583	1.5860	0.0951	1.9917	0.1231
inconsistency error $u(\mathbf{x}) + v(\mathbf{x} + u(\mathbf{x}))$ on contours in S						
Algorithm 1	1.6165	0.0524	1.9630	0.0459	2.3577	0.0770
Scheme (2)	4.7525	0.1036	6.2510	0.1292	6.3180	0.1843

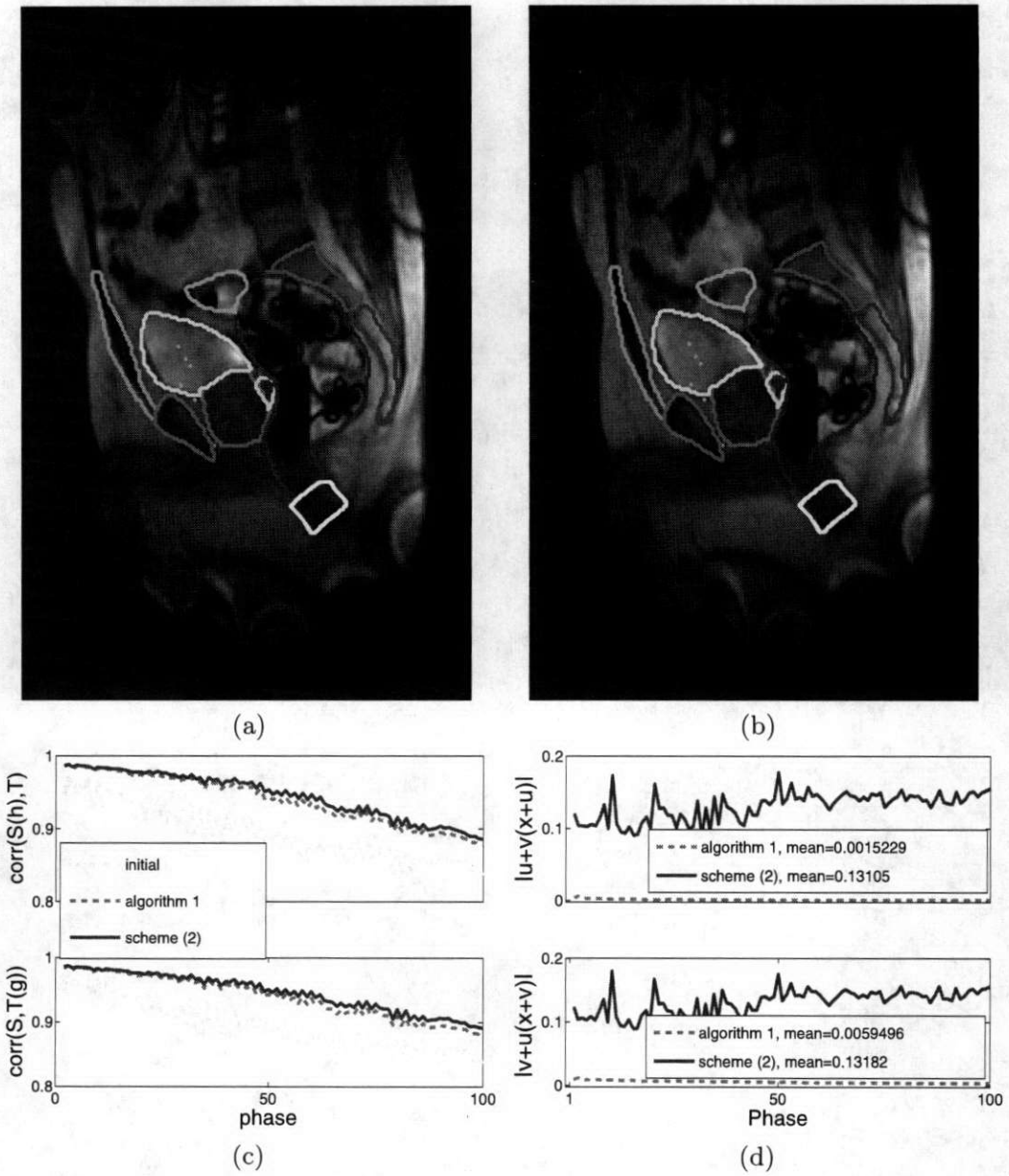


Fig. 2. Experiment on 3D prostate MRI data with 100 2D phases. (a) Source image(the $1st$ phase) with contours of ROIs superimposed, (b) the $21st$ phase with automatic re-contouring results by Algorithm 1 superimposed, (c) plots of correlation $Corr(S(h), T)$, and $Corr(S, T(g))$ between the $1st$ phase and each of other 99 phases, (d) plots of norms $||u(x) + v(x + u(x))||$ and $||v(x) + u(x + v(x))||$ for deformations between the $1st$ phase and each of other 99 phases, (c,d) are based on parameters $\lambda = 20, dt = .1$ for 200 iterations.

focusing on prostate area, where ROIs have large internal motions. The source volume S is the first phase, and the boundaries of ROIs in S are delineated by contours and superimposed in Fig 2(a), and the other 99 phases are targets, and

 3195

methods Scheme (2) and Algorithm 1 are applied to find deformations between the $1st$ phase and each of the other 99 phases. The auto re-contouring, that register the contours in S into the other 99 phases, is achieved by applying the deformations on these contours. For demonstration purpose, one of the target image T, the $21st$ phase, is shown in Fig 2(b), and the auto re-contouring result by Algorithm 1 is superimposed on it.

Comparison on the convergence and inverse inconsistency by the first two methods are shown in Fig 2(c,d). Fig 2(c) compares CC between deformed images and target images by these two models. Both Scheme (2) and Algorithm 1 improve the initial CC between S and each of other 99 phases to a similar level. However, from Fig 2(d) we can observe that the norm of inverse inconsistency error by scheme (2) is much higher in average than that of Algorithm 1.

The quantitative comparison on inverse inconsistency errors between the $1st$ phase and the 21st phase for this experiment is listed in Table 2 for demonstration. Beside comparing the error fields on Ω, we also evaluate error $v(x)+u(x+v)$ at points of all given contours on S. By comparing the corresponding components and norms of the inverse inconsistency error fields in Table 2, we find that errors generated by proposed algorithm is much lower than that by scheme (2), this indicates that the point correspondence and automatic re-contouring results are more accurate.

4 Conclusion

In this work, we proposed a coupled energy minimization method with inverse consistent constraints for deformable image registration. The proposed model controls the inverse inconsistency errors in a negligibly low level, therefore, it provides a better correspondence for the ROIs in source and target images. This makes the auto re-contouring results with data involved in the course of radiation therapy much more accurate.

The dissimilarity measure used in this work is the negative log-likelihood of the residual image between the deformed source and target. This dissimilarity measure is able to accommodate certain variability in the matching. Hence, the model is more robust to noise than SSD, moreover, it is less sensitive to the choice of the parameter that balances the smoothness of the deformation field and goodness of matching.

Acknowledgments

Authors would like to thank Dr. Jim Dempsey and Dr. Anneyuko I. Saito from Department of Radiation Oncology at University of Florida and Viewray Inc. for providing the prostate data. The work was partially supported by NIH R01NS052831-01A1.

References

1. Bajcsy, R., Kovacic, S.: Multiresolution elastic matching. Computer Vision, Graphics, and Image Processing 46(1), 1–21 (1989)
2. Beg, M.F., Khan, A.: Symmetric data attachment terms for large deformation image registration. MedImg 26(9), 1179–1189 (2007)
3. Christensen, G., Johnson, H.: Consistent image registration. IEEE Transactions on Medical Imaging 20, 568–582 (2001)
4. Collins, D.L., Evans, A.C.: Animal: Validation and applications of nonlinear registration-based segmentation. IJPRAI 11(8), 1271–1294 (1997)
5. Schreibmann, E., Xing, L.: Narrow band deformable registration of prostate magnetic resonance imaging, magnetic resonance spectroscopic imaging, and computed tomography studies. Int. J. Radiat. Oncol. Biol. Phys. 62, 595–605 (2005)
6. Hermosillo, G., Hotel, C.C., Faugeras, O.: Variational methods for multimodal image matching. Int. J. Computer Vision 50(3), 329–343 (2002)
7. Hill, D., Batchelor, P., Holden, M., Hawkes, D.: Topical review: Medical image registration. Physics in Medicine and Biology 46, 1–45 (2001)
8. Jian, B., Vemuri, B., Marroquín, J.: Robust nonrigid multimodal image registration using local frequency maps. In: 19th International Conference on Information Processing in Medical Imaging, pp. 504–515 (2005)
9. Weickert, J., Romeny, B., Viergever, M.: Efficient and reliable schemes for nonlinear diffusion filtering. IEEE Trans. on Img. Proc. 7(3), 398–410 (1998)
10. Brock, K.K., Sharpe, M.B., Dawson, L.A., Kim, S.M., Jaffray, D.A.: Accuracy of finite element model-based multi-organ deformable image registration. Med. Phys. 32, 1647–1659 (2005)
11. Keall, P.J., Siebers, J.V., Joshi, S., Mohan, R.: Monte carlo as a four-dimensional radiotherapy treatment-planning tool to account for respiratory motion. Phys. Med. Biol. 49, 3639–3648 (2004)
12. Leow, A.D., Huang, S.C., Geng, A., Becker, J., Davis, S., Toga, A., Thompson, P.: Inverse consistent mapping in 3d deformable image registration: Its construction and statistical properties. In: Christensen, G.E., Sonka, M. (eds.) IPMI 2005. LNCS, vol. 3565, pp. 493–503. Springer, Heidelberg (2005)
13. Lu, T., Neittaanmki, P., Tai, X.-C.: A parallel splitting-up method for partial differential equations and its application to navier-stokes equations. RAIRO Math. Model. and Numer. Anal. 26(6), 673–708 (1992)
14. Lu, W., Chen, M., Olivera, G., Ruchala, K., Mackie, T.: Fast free-form deformable registration via calculus of variations. Physics in Medicine and Biology 49, 3067–3087 (2004)
15. Lu, W., Olivera, G., Mackie, T.R.: Motion-encoded dose calculation through fluence/sinogram modification. Med. Phys. 32, 118–127 (2005)
16. Lu, W., Olivera, G.H., Chen, Q., Chen, M., Ruchala, K.: Automatic re-contouring in 4d radiotherapy. Physics in Medicine and Biology 51, 1077–1099 (2006)
17. Coselmon, M.M., Balter, J.M., McShan, D.L., Kessler, M.L.: Mutual information based ct registration of the lung at exhale and inhale breathing states using thin-plate splines. Med. Phys. 31, 2942–2948 (2004)
18. Maes, F., Collignon, A., Vandermeulen, D., Marchal, G., Suetens, P.: Multimodality image registration maximization of mutual information. In: MMBIA 1996: Proceedings of the 1996 Workshop on Mathematical Methods in Biomedical Image Analysis (MMBIA 1996), Washington, DC, USA, p. 14. IEEE Computer Society Press, Los Alamitos (1996)

19. Rogelj, P., Kovacic, S.: Similarity measures for non-rigid registration. In: Proc. SPIE, vol. 4322, pp. 569–578 (2001)
20. Thirion, J.: Image matching as a diffusion process: an analogy with maxwell's demons. Medical image analysis 2(3), 243–260 (1998)
21. Viola, P.A., Wells III, W.M.: Alignment by maximization of mutual information. In: ICCV, pp. 16–23 (1995)
22. Pluim, J.P.W., Maintz, J.B.A., Viergever, M.A.: Mutual-information-based registration of medical images: A survey. IEEE Trans. Med. Imaging 22, 986–1004 (2003)
23. Wang, H., Dong, L., O'Daniel, J., Mohan, R., Garden, A., Ang, K., Kuban, D., Bonnen, M., Chang, J., Cheung, R.: Validation of an accelerated 'demons' algorithm for deformable image registration in radiation therapy. Phys. Med. Biol. 50(12), 2887–2905 (2005)
24. Zitova, B., Flusser, J.: Image registration methods: A survey. Image Vis. comput. 21, 977–1000 (2003)

Numerical Recipes in C++

The Art of Scientific Computing

Second Edition

William H. Press

Los Alamos National Laboratory

Saul A. Teukolsky

Department of Physics, Cornell University

William T. Vetterling

Polaroid Corporation

Brian P. Flannery

EXXON Research and Engineering Company

CAMBRIDGE
UNIVERSITY PRESS

Chapter 13. Fourier and Spectral Applications

13.0 Introduction

Fourier methods have revolutionized fields of science and engineering, from radio astronomy to medical imaging, from seismology to spectroscopy. In this chapter, we present some of the basic applications of Fourier and spectral methods that have made these revolutions possible.

Say the word "Fourier" to a numericist, and the response, as if by Pavlovian conditioning, will likely be "FFT." Indeed, the wide application of Fourier methods must be credited principally to the existence of the fast Fourier transform. Better mousetraps stand aside: If you speed up *any* nontrivial algorithm by a factor of a million or so, the world will beat a path towards finding useful applications for it. The most direct applications of the FFT are to the convolution or deconvolution of data (§13.1), correlation and autocorrelation (§13.2), optimal filtering (§13.3), power spectrum estimation (§13.4), and the computation of Fourier integrals (§13.9).

As important as they are, however, FFT methods are not the be-all and end-all of spectral analysis. Section 13.5 is a brief introduction to the field of time-domain digital filters. In the spectral domain, one limitation of the FFT is that it always represents a function's Fourier transform as a polynomial in $z = \exp(2\pi i f \Delta)$ (cf. equation 12.1.7). Sometimes, processes have spectra whose shapes are not well represented by this form. An alternative form, which allows the spectrum to have poles in z, is used in the techniques of linear prediction (§13.6) and maximum entropy spectral estimation (§13.7).

Another significant limitation of all FFT methods is that they require the input data to be sampled at evenly spaced intervals. For irregularly or incompletely sampled data, other (albeit slower) methods are available, as discussed in §13.8.

So-called wavelet methods inhabit a representation of function space that is neither in the temporal, nor in the spectral, domain, but rather something in-between. Section 13.10 is an introduction to this subject. Finally §13.11 is an excursion into numerical use of the Fourier sampling theorem.

13.1 Convolution and Deconvolution Using the FFT

We have defined the *convolution* of two functions for the continuous case in equation (12.0.8), and have given the *convolution theorem* as equation (12.0.9). The theorem says that the Fourier transform of the convolution of two functions is equal to the product of their individual Fourier transforms. Now, we want to deal with the discrete case. We will mention first the context in which convolution is a useful procedure, and then discuss how to compute it efficiently using the FFT.

The convolution of two functions $r(t)$ and $s(t)$, denoted $r * s$, is mathematically equal to their convolution in the opposite order, $s * r$. Nevertheless, in most applications the two functions have quite different meanings and characters. One of the functions, say s, is typically a signal or data stream, which goes on indefinitely in time (or in whatever the appropriate independent variable may be). The other function r is a "response function," typically a peaked function that falls to zero in both directions from its maximum. The effect of convolution is to smear the signal $s(t)$ in time according to the recipe provided by the response function $r(t)$, as shown in Figure 13.1.1. In particular, a spike or delta-function of unit area in s which occurs at some time t_0 is supposed to be smeared into the shape of the response function itself, but translated from time 0 to time t_0 as $r(t - t_0)$.

In the discrete case, the signal $s(t)$ is represented by its sampled values at equal time intervals s_j. The response function is also a discrete set of numbers r_k, with the following interpretation: r_0 tells what multiple of the input signal in one channel (one particular value of j) is copied into the identical output channel (same value of j); r_1 tells what multiple of input signal in channel j is additionally copied into output channel $j + 1$; r_{-1} tells the multiple that is copied into channel $j - 1$; and so on for both positive and negative values of k in r_k. Figure 13.1.2 illustrates the situation.

Example: a response function with $r_0 = 1$ and all other r_k's equal to zero is just the identity filter: convolution of a signal with this response function gives identically the signal. Another example is the response function with $r_{14} = 1.5$ and all other r_k's equal to zero. This produces convolved output that is the input signal multiplied by 1.5 and delayed by 14 sample intervals.

Evidently, we have just described in words the following definition of discrete convolution with a response function of finite duration M:

$$(r * s)_j \equiv \sum_{k=-M/2+1}^{M/2} s_{j-k}\, r_k \tag{13.1.1}$$

If a discrete response function is nonzero only in some range $-M/2 < k \le M/2$, where M is a sufficiently large even integer, then the response function is called a *finite impulse response (FIR)*, and its *duration* is M. (Notice that we are defining M as the number of nonzero *values* of r_k; these values span a time interval of $M - 1$ sampling times.) In most practical circumstances the case of finite M is the case of interest, either because the response really has a finite duration, or because we choose to truncate it at some point and approximate it by a finite-duration response function.

The *discrete convolution theorem* is this: If a signal s_j is *periodic* with period N, so that it is completely determined by the N values $s_0, \ldots, s_{N-1}$, then its

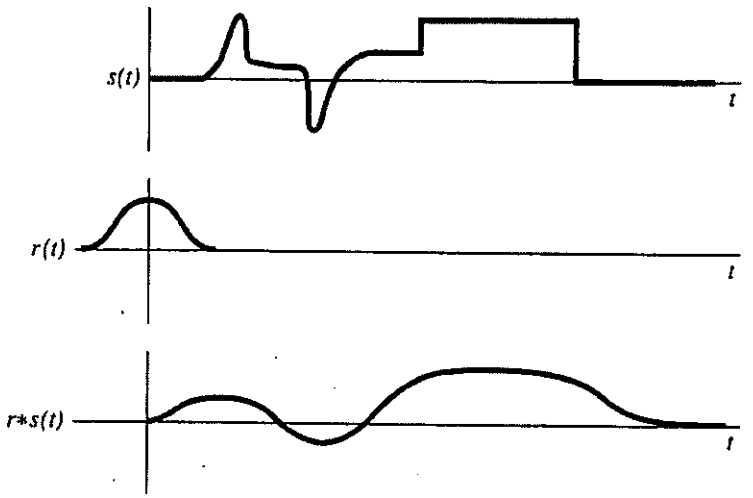


Figure 13.1.1. Example of the convolution of two functions. A signal $s(t)$ is convolved with a response function $r(t)$. Since the response function is broader than some features in the original signal, these are "washed out" in the convolution. In the absence of any additional noise, the process can be reversed by deconvolution.

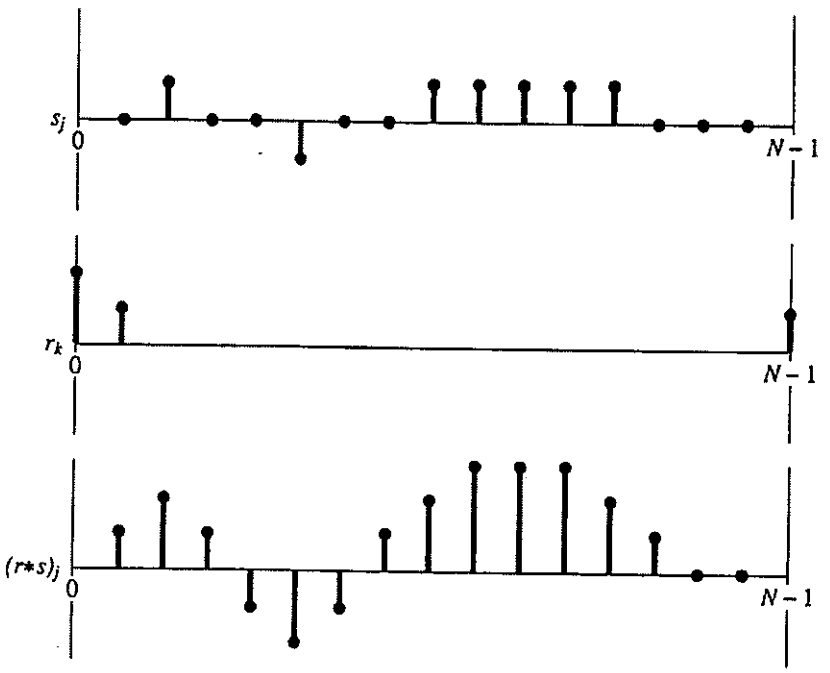


Figure 13.1.2. Convolution of discretely sampled functions. Note how the response function for negative times is wrapped around and stored at the extreme right end of the array r_k.

discrete convolution with a response function *of finite duration* N is a member of the discrete Fourier transform pair,

$$\sum_{k=-N/2+1}^{N/2} s_{j-k}\, r_k \quad \Longleftrightarrow \quad S_n R_n \tag{13.1.2}$$

Here S_n, $(n = 0, \ldots, N-1)$ is the discrete Fourier transform of the values s_j, $(j = 0, \ldots, N-1)$, while R_n, $(n = 0, \ldots, N-1)$ is the discrete Fourier transform of the values r_k, $(k = 0, \ldots, N-1)$. These values of r_k are the same ones as for the range $k = -N/2+1, \ldots, N/2$, but in wrap-around order, exactly as was described at the end of §12.2.

Treatment of End Effects by Zero Padding

The discrete convolution theorem presumes a set of two circumstances that are not universal. First, it assumes that the input signal is periodic, whereas real data often either go forever without repetition or else consist of one nonperiodic stretch of finite length. Second, the convolution theorem takes the duration of the response to be the same as the period of the data; they are both N. We need to work around these two constraints.

The second is very straightforward. Almost always, one is interested in a response function whose duration M is much shorter than the length of the data set N. In this case, you simply extend the response function to length N by padding it with zeros, i.e., define $r_k = 0$ for $M/2 \le k \le N/2$ and also for $-N/2+1 \le k \le -M/2+1$. Dealing with the first constraint is more challenging. Since the convolution theorem rashly assumes that the data are periodic, it will falsely "pollute" the first output channel $(r * s)_0$ with some wrapped-around data from the far end of the data stream s_{N-1}, s_{N-2}, etc. (See Figure 13.1.3.) So, we need to set up a buffer zone of zero-padded values at the end of the s_j vector, in order to make this pollution zero. How many zero values do we need in this buffer? Exactly as many as the most negative index for which the response function is nonzero. For example, if r_{-3} is nonzero, while $r_{-4}, r_{-5}, \ldots$ are all zero, then we need three zero pads at the end of the data: $s_{N-3} = s_{N-2} = s_{N-1} = 0$. These zeros will protect the first output channel $(r * s)_0$ from wrap-around pollution. It should be obvious that the second output channel $(r * s)_1$ and subsequent ones will also be protected by these same zeros. Let K denote the number of padding zeros, so that the last actual input data point is s_{N-K-1}.

What now about pollution of the very *last* output channel? Since the data now end with s_{N-K-1}, the last output channel of interest is $(r * s)_{N-K-1}$. This channel can be polluted by wrap-around from input channel s_0 unless the number K is also large enough to take care of the most positive index k for which the response function r_k is nonzero. For example, if r_0 through r_6 are nonzero, while $r_7, r_8 \ldots$ are all zero, then we need at least $K = 6$ padding zeros at the end of the data: $s_{N-6} = \ldots = s_{N-1} = 0$.

To summarize — we need to pad the data with a number of zeros *on one end* equal to the maximum positive duration *or* maximum negative duration of the response function, *whichever is larger*. (For a symmetric response function of duration M, you will need only $M/2$ zero pads.) Combining this operation with the

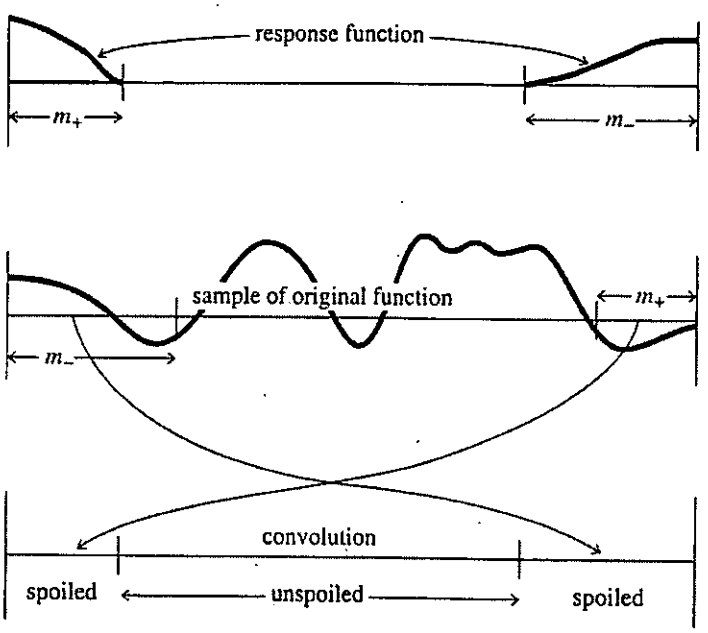


Figure 13.1.3. The wrap-around problem in convolving finite segments of a function. Not only must the response function wrap be viewed as cyclic, but so must the sampled original function. Therefore a portion at each end of the original function is erroneously wrapped around by convolution with the response function.

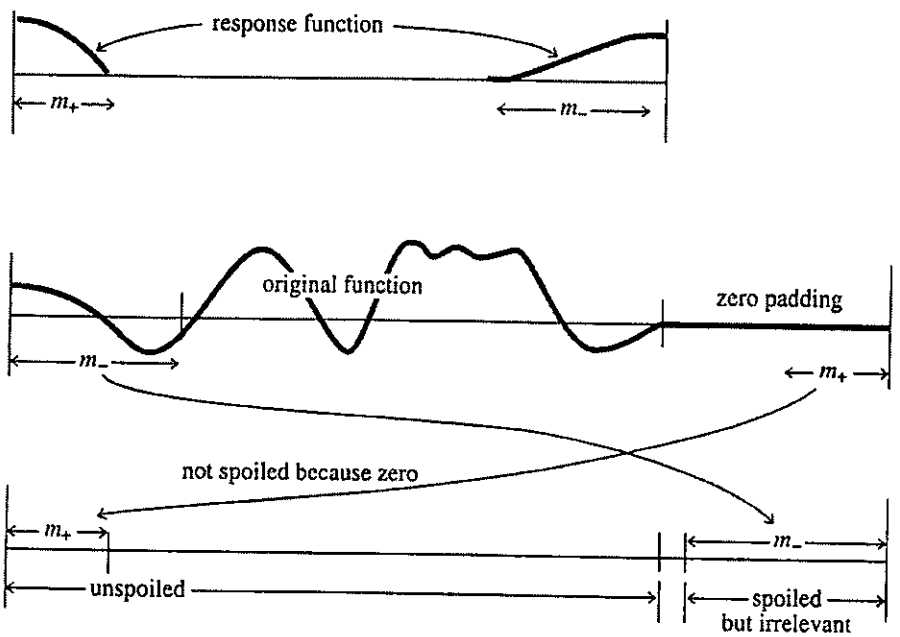


Figure 13.1.4. Zero padding as solution to the wrap-around problem. The original function is extended by zeros, serving a dual purpose: When the zeros wrap around, they do not disturb the true convolution; and while the original function wraps around onto the zero region, that region can be discarded.

padding of the response r_k described above, we effectively insulate the data from artifacts of undesired periodicity. Figure 13.1.4 illustrates matters.

Use of FFT for Convolution

The data, complete with zero padding, are now a set of real numbers s_j, $j = 0, \ldots, N-1$, and the response function is zero padded out to duration N and arranged in wrap-around order. (Generally this means that a large contiguous section of the r_k's, in the middle of that array, is zero, with nonzero values clustered at the two extreme ends of the array.) You now compute the discrete convolution as follows: Use the FFT algorithm to compute the discrete Fourier transform of s and of r. Multiply the two transforms together component by component, remembering that the transforms consist of complex numbers. Then use the FFT algorithm to take the inverse discrete Fourier transform of the products. The answer is the convolution $r * s$.

What about *deconvolution*? Deconvolution is the process of *undoing* the smearing in a data set that has occurred under the influence of a known response function, for example, because of the known effect of a less-than-perfect measuring apparatus. The defining equation of deconvolution is the same as that for convolution, namely (13.1.1), except now the left-hand side is taken to be known, and (13.1.1) is to be considered as a set of N linear equations for the unknown quantities s_j. Solving these simultaneous linear equations in the time domain of (13.1.1) is unrealistic in most cases, but the FFT renders the problem almost trivial. Instead of multiplying the transform of the signal and response to get the transform of the convolution, we just divide the transform of the (known) convolution by the transform of the response to get the transform of the deconvolved signal.

This procedure can go wrong *mathematically* if the transform of the response function is exactly zero for some value R_n, so that we can't divide by it. This indicates that the original convolution has truly lost all information at that one frequency, so that a reconstruction of that frequency component is not possible. You should be aware, however, that apart from mathematical problems, the process of deconvolution has other practical shortcomings. The process is generally quite sensitive to noise in the input data, and to the accuracy to which the response function r_k is known. Perfectly reasonable attempts at deconvolution can sometimes produce nonsense for these reasons. In such cases you may want to make use of the additional process of *optimal filtering*, which is discussed in §13.3.

Here is our routine for convolution and deconvolution, using the FFT as implemented in `four1` of §12.2. Since the data and response functions are real, not complex, both of their transforms could be taken simultaneously using `twofft`. However, since `data` and `respns` often have very different magnitudes, we instead make separate calls to `realft` to minimize roundoff. The data are assumed to be stored in a `double` array `data[0..n-1]`, with `n` an integer power of two. The response function is assumed to be stored in wrap-around order in a `double` array `respns[0..m-1]`. The value of `m` can be any *odd* integer less than or equal to `n`, since the first thing the program does is to recopy the response function into the appropriate wrap-around order in an array of length `n`. The answer is provided in `ans`, which is also used as working space.

```
#include "nr.h"

void NR::convlv(Vec_I_DP &data, Vec_I_DP &respns, const int isign,
    Vec_O_DP &ans)
```

Convolves or deconvolves a real data set `data[0..n-1]` (including any user-supplied zero padding) with a response function `respns[0..m-1]`, where `m` is an odd integer $\leq$ `n`. The response function must be stored in wrap-around order: the first half of the array `respns` contains the impulse response function at positive times, while the second half of the array contains the impulse response function at negative times, counting down from the highest element `respns[m-1]`. On input `isign` is $+1$ for convolution, -1 for deconvolution. The answer is returned in `ans[0..n-1]`. `n` MUST be an integer power of two.

```
{
    int i,no2;
    DP mag2,tmp;

    int n=data.size();
    int m=respns.size();
    Vec_DP temp(n);
    temp[0]=respns[0];
    for (i=1;i<(m+1)/2;i++) {                    Put respns in array of length n.
        temp[i]=respns[i];
        temp[n-i]=respns[m-i];
    }
    for (i=(m+1)/2;i<n-(m-1)/2;i++)              Pad with zeros.
        temp[i]=0.0;
    for (i=0;i<n;i++)
        ans[i]=data[i];
    realft(ans,1);                               FFT both arrays.
    realft(temp,1);
    no2=n>>1;
    if (isign == 1) {
        for (i=2;i<n;i+=2) {                     Multiply FFTs to convolve.
            tmp=ans[i];
            ans[i]=(ans[i]*temp[i]-ans[i+1]*temp[i+1])/no2;
            ans[i+1]=(ans[i+1]*temp[i]+tmp*temp[i+1])/no2;
        }
        ans[0]=ans[0]*temp[0]/no2;
        ans[1]=ans[1]*temp[1]/no2;
    } else if (isign == -1) {
        for (i=2;i<n;i+=2) {                     Divide FFTs to deconvolve.
            if ((mag2=SQR(temp[i])+SQR(temp[i+1])) == 0.0)
                nrerror("Deconvolving at response zero in convlv");
            tmp=ans[i];
            ans[i]=(ans[i]*temp[i]+ans[i+1]*temp[i+1])/mag2/no2;
            ans[i+1]=(ans[i+1]*temp[i]-tmp*temp[i+1])/mag2/no2;
        }
        if (temp[0] == 0.0 || temp[1] == 0.0)
            nrerror("Deconvolving at response zero in convlv");
        ans[0]=ans[0]/temp[0]/no2;
        ans[1]=ans[1]/temp[1]/no2;
    } else nrerror("No meaning for isign in convlv");
    realft(ans,-1);                              Inverse transform back to time domain.
}
```

Convolving or Deconvolving Very Large Data Sets

If your data set is so long that you do not want to fit it into memory all at once, then you must break it up into sections and convolve each section separately. Now, however, the treatment of end effects is a bit different. You have to worry not only about spurious wrap-around effects, but also about the fact that the ends of

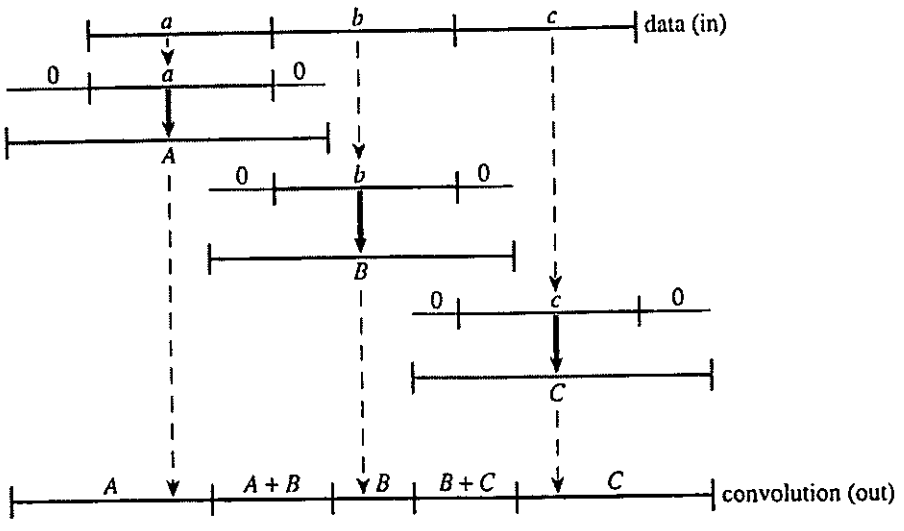


Figure 13.1.5. The overlap-add method for convolving a response with a very long signal. The signal data is broken up into smaller pieces. Each is zero padded at both ends and convolved (denoted by bold arrows in the figure). Finally the pieces are added back together, including the overlapping regions formed by the zero pads.

each section of data *should* have been influenced by data at the nearby ends of the immediately preceding and following sections of data, but were not so influenced since only one section of data is in the machine at a time.

There are two, related, standard solutions to this problem. Both are fairly obvious, so with a few words of description here, you ought to be able to implement them for yourself. The first solution is called the *overlap-save method*. In this technique you pad only the very beginning of the data with enough zeros to avoid wrap-around pollution. After this initial padding, you forget about zero padding altogether. Bring in a section of data and convolve or deconvolve it. Then throw out the points at each end that are polluted by wrap-around end effects. Output only the remaining good points in the middle. Now bring in the next section of data, but not all new data. The first points in each new section overlap the last points from the preceding section of data. The sections must be overlapped sufficiently so that the polluted output points at the end of one section are recomputed as the first of the unpolluted output points from the subsequent section. With a bit of thought you can easily determine how many points to overlap and save.

The second solution, called the *overlap-add method*, is illustrated in Figure 13.1.5. Here you *don't* overlap the input data. Each section of data is disjoint from the others and is used exactly once. However, you carefully zero-pad it at both ends so that there is no wrap-around ambiguity in the output convolution or deconvolution. Now you overlap *and add* these sections of output. Thus, an output point near the end of one section will have the response due to the input points at the beginning of the next section of data properly added in to it, and likewise for an output point near the beginning of a section, *mutatis mutandis*.

Even when computer memory is available, there is some slight gain in computing speed in segmenting a long data set, since the FFTs' $N \log_2 N$ is slightly slower than linear in N. However, the log term is so slowly varying that you will often be much

happier to avoid the bookkeeping complexities of the overlap-add or overlap-save methods: If it is practical to do so, just cram the whole data set into memory and FFT away. Then you will have more time for the finer things in life, some of which are described in succeeding sections of this chapter.

CITED REFERENCES AND FURTHER READING:

Nussbaumer, H.J. 1982, *Fast Fourier Transform and Convolution Algorithms* (New York: Springer-Verlag).

Elliott, D.F., and Rao, K.R. 1982, *Fast Transforms: Algorithms, Analyses, Applications* (New York: Academic Press).

Brigham, E.O. 1974, *The Fast Fourier Transform* (Englewood Cliffs, NJ: Prentice-Hall), Chapter 13.

13.2 Correlation and Autocorrelation Using the FFT

Correlation is the close mathematical cousin of convolution. It is in some ways simpler, however, because the two functions that go into a correlation are not as conceptually distinct as were the data and response functions that entered into convolution. Rather, in correlation, the functions are represented by different, but generally similar, data sets. We investigate their "correlation," by comparing them both directly superposed, and with one of them shifted left or right.

We have already defined in equation (12.0.10) the correlation between two continuous functions $g(t)$ and $h(t)$, which is denoted $\mathrm{Corr}(g,h)$, and is a function of *lag* t. We will occasionally show this time dependence explicitly, with the rather awkward notation $\mathrm{Corr}(g,h)(t)$. The correlation will be large at some value of t if the first function (g) is a close copy of the second (h) but lags it in time by t, i.e., if the first function is shifted to the right of the second. Likewise, the correlation will be large for some negative value of t if the first function *leads* the second, i.e., is shifted to the left of the second. The relation that holds when the two functions are interchanged is

$$\mathrm{Corr}(g,h)(t) = \mathrm{Corr}(h,g)(-t) \tag{13.2.1}$$

The discrete correlation of two sampled functions g_k and h_k, each periodic with period N, is defined by

$$\mathrm{Corr}(g,h)_j \equiv \sum_{k=0}^{N-1} g_{j+k} h_k \tag{13.2.2}$$

The *discrete correlation theorem* says that this discrete correlation of two real functions g and h is one member of the discrete Fourier transform pair

$$\mathrm{Corr}(g,h)_j \Longleftrightarrow G_k H_k^* \tag{13.2.3}$$

MARCHING CUBES: A HIGH RESOLUTION 3D SURFACE CONSTRUCTION ALGORITHM

William E. Lorensen
Harvey E. Cline

General Electric Company
Corporate Research and Development
Schenectady, New York 12301

Abstract

We present a new algorithm, called *marching cubes*, that creates triangle models of constant density surfaces from 3D medical data. Using a divide-and-conquer approach to generate inter-slice connectivity, we create a case table that defines triangle topology. The algorithm processes the 3D medical data in scan-line order and calculates triangle vertices using linear interpolation. We find the gradient of the original data, normalize it, and use it as a basis for shading the models. The detail in images produced from the generated surface models is the result of maintaining the inter-slice connectivity, surface data, and gradient information present in the original 3D data. Results from computed tomography (CT), magnetic resonance (MR), and single-photon emission computed tomography (SPECT) illustrate the quality and functionality of *marching cubes*. We also discuss improvements that decrease processing time and add solid modeling capabilities.

CR Categories: 3.3, 3.5

Additional Keywords: computer graphics, medical imaging, surface reconstruction

1. INTRODUCTION.

Three-dimensional surfaces of the anatomy offer a valuable medical tool. Images of these surfaces, constructed from multiple 2D slices of computed tomography (CT), magnetic resonance (MR), and single-photon emission computed tomography (SPECT), help physicians to understand the complex anatomy present in the slices. Interpretation of 2D medical images requires special training, and although radiologists have these skills, they must often communicate their interpretations to the referring physicians, who sometimes have difficulty visualizing the 3D anatomy.

Researchers have reported the application of 3D medical images in a variety of areas. The visualization of complex acetabular fractures [6], craniofacial abnormalities [17,18], and intracranial structure [13] illustrate 3D's potential for the study of complex bone structures. Applications in radiation therapy [27,11] and surgical planning [4,5,31] show interactive 3D techniques combined with 3D surface images. Cardiac applications include artery visualization [2,16] and non-graphic modeling applications to calculate surface area and volume [21].

Existing 3D algorithms lack detail and sometimes introduce artifacts. We present a new, high-resolution 3D surface construction algorithm that produces models with unprecedented detail. This new algorithm, called *marching cubes*, creates a polygonal representation of constant density surfaces from a 3D array of data. The resulting model can be displayed with conventional graphics-rendering algorithms implemented in software or hardware.

After describing the information flow for 3D medical applications, we describe related work and discuss the drawbacks of that work. Then we describe the algorithm as well as efficiency and functional enhancements, followed by case studies using three different medical imaging techniques to illustrate the new algorithm's capabilities.

2. INFORMATION FLOW FOR 3D MEDICAL ALGORITHMS.

Medical applications of 3D consist of four steps (Figure 1). Although one can combine the last three steps into one algorithm, we logically decompose the process as follows:

1. *Data acquisition.*
   This first step, performed by the medical imaging hardware, samples some property in a patient and produces multiple 2D slices of information. The data sampled depends on the data acquisition technique.

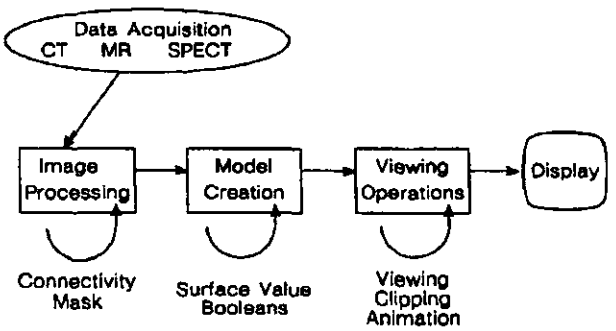


Figure 1. 3D Medical Information Flow.

ACM-0-89791-227-6/87/007/0163 $00.75

X-ray computed tomography (CT) measures the spatially varying X-ray attenuation coefficient [3]. CT images show internal structure. For 3D applications, CT is frequently used to look at bone structure, although we have had success visualizing soft tissue.

Magnetic resonance (MR) measures three physical properties [20]. One property is the distribution of "mobile" hydrogen nuclei and shows overall structure within the slices. The other two properties measure relaxation times of the nuclei. MR, a recent technique, shows excellent contrast between a variety of soft tissues. However, the variety of surfaces presents a challenge to 3D surface construction and requires techniques for selective surface extraction and display.

A third acquisition technique, single-photon emission computed tomography (SPECT) measures the emission of gamma rays [24]. The source of these rays is a radioisotope distributed within the body. In addition to structure, SPECT can show the presence of blood in structures with a much lower dose than that required by CT.

2. *Image processing.*
Some algorithms use image processing techniques to find structures within the 3D data [1,32,30,29] or to filter the original data. MR data, in particular, needs image processing to select appropriate structure.

3. *Surface construction.*
Surface construction, the topic of this paper, involves the creation of a surface model from the 3D data. The model usually consists of 3D volume elements (voxels) or polygons. Users select the desired surface by specifying a density value. This step can also include the creation of cut or capped surfaces.

4. *Display.*
Having created the surface, the final step displays that surface using display techniques that include ray casting, depth shading, and color shading.

3. RELATED WORK.

There are several approaches to the 3D surface generation problem. An early technique [23] starts with contours of the surface to be constructed and connects contours on consecutive slices with triangles. Unfortunately, if more than one contour of surface exists on a slice, ambiguities arise when determining which contours to connect [14]. Interactive intervention by the user can overcome some of these ambiguities [8]; however, in a clinical environment, user interaction should be kept to a minimum.

Another approach, developed by G. Herman and colleagues [19] creates surfaces from cuberilles. A cuberille is "dissection of space into equal cubes (called voxels) by three orthogonal sets of parallel planes [7]." Although there are many ways to display a cuberille model, the most realistic images result when the gradient, calculated from cuberilles in a neighborhood, is used to find the shade of a point on the model [15]. Meagher [25] uses an octree representation to compress the storage of the 3D data, allowing rapid manipulation and display of voxels.

Farrell [12] uses ray casting to find the 3D surface, but rather than shade the image with a gray scale, uses hue lightness to display the surface. In another ray casting method, Hohne [22], after locating the surface along a ray, calculates the gradient along the surface and uses this gradient, scaled by an "appropriate" value, to generate gray scales for the image.

A different approach, used at the Mayo Clinic [26], displays the density volume rather than the surface. This method produces, in effect, a conventional shadow graph that can be viewed from arbitrary angles. Motion enhances the three-dimensional effect obtained using the volume model.

Each of these techniques for surface construction and display suffer shortcomings because they throw away useful information in the original data. The connected contour algorithms throw away the inter-slice connectivity that exists in the original data. The cuberille approach, using thresholding to represent the surface as blocks in 3D space, attempts to recover shading information from the blocks. The ray casting methods either use depth shading alone, or try to approximate shading with an unnormalized gradient. Since they display all values and not just those visible from a given point of view, volume models rely on motion to produce a three-dimensional sensation.

Our approach uses information from the original 3D data to derive inter-slice connectivity, surface location, and surface gradient. The resulting triangle model can be displayed on conventional graphics display systems using standard rendering algorithms.

4. MARCHING CUBES ALGORITHM.

There are two primary steps in our approach to the surface construction problem. First, we locate the surface corresponding to a user-specified value and create triangles. Then, to ensure a quality image of the surface, we calculate the normals to the surface at each vertex of each triangle.

Marching cubes uses a divide-and-conquer approach to locate the surface in a logical *cube* created from eight pixels; four each from two adjacent slices (Figure 2).

The algorithm determines how the surface intersects this cube, then moves (or *marchs*) to the next cube. To find the surface intersection in a cube, we assign a one to a cube's vertex if the data value at that vertex exceeds (or equals) the value of the surface we are constructing. These vertices are inside (or on) the surface. Cube vertices with values below the surface receive a zero and are outside the surface. The surface intersects those cube edges where one vertex is outside the surface (one) and the other is inside the surface (zero). With this assumption, we determine the topology of the surface within a cube, finding the location of the intersection later.

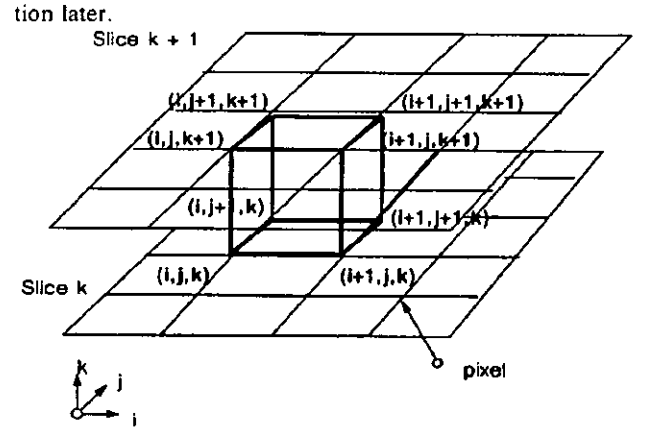


Figure 2. Marching Cube.

Since there are eight vertices in each cube and two states, inside and outside, there are only $2^8 = 256$ ways a surface can intersect the cube. By enumerating these 256 cases, we create a table to look up surface-edge intersections, given the labeling of a cubes vertices. The table contains the edges intersected for each case.

Triangulating the 256 cases is possible but tedious and error-prone. Two different symmetries of the cube reduce the problem from 256 cases to 14 patterns. First, the topology of the triangulated surface is unchanged if the relationship of the surface values to the cubes is reversed. Complementary cases, where vertices greater than the surface value are interchanged with those less than the value, are equivalent. Thus, only cases with zero to four vertices greater than the surface value need be considered, reducing the number of cases to 128. Using the second symmetry property, rotational symmetry, we reduced the problem to 14 patterns by inspection. Figure 3 shows the triangulation for the 14 patterns.

The simplest pattern, 0, occurs if all vertex values are above (or below) the selected value and produces no triangles. The next pattern, 1, occurs if the surface separates on vertex from the other seven, resulting in one triangle defined by the three edge intersections. Other patterns produce multiple triangles. Permutation of these 14 basic patterns using complementary and rotational symmetry produces the 256 cases.

We create an index for each case, based on the state of the vertex. Using the vertex numbering in Figure 4, the eight bit index contains one bit for each vertex.

This index serves as a pointer into an edge table that gives all edge intersections for a given cube configuration.

Using the index to tell which edge the surface intersects, we can interpolate the surface intersection along the edge. We use linear interpolation, but have experimented with higher degree interpolations. Since the algorithm produces at least one and as many as four triangles per cube, the higher degree surfaces show little improvement over linear interpolation.

The final step in *marching cubes* calculates a unit normal for each triangle vertex. The rendering algorithms use this normal to produce Gouraud-shaded images. A surface of constant density has a zero gradient component along the surface tangential direction; consequently, the direction of the gradient vector, $\vec{g}$, is normal to the surface. We can use this fact to determine surface normal vector, $\vec{n}$, if the magnitude of the gradient, $|\vec{g}|$, is nonzero. Fortunately, at the surface of interest between two tissue types of different densities, the gradient vector is nonzero. The gradient vector, $\vec{g}$, is the derivative of the density function

$$\vec{g}(x,y,z) = \nabla \vec{f}(x,y,z). \tag{1}$$

To estimate the gradient vector at the surface of interest, we first estimate the gradient vectors at the cube vertices and linearly interpolate the gradient at the point of intersection. The gradient at cube vertex (i,j,k), is estimated using central differences along the three coordinate axes by:

$$G_x(i,j,k) = \frac{D(i+1,j,k) - D(i-1,j,k)}{\Delta x} \tag{2}$$

$$G_y(i,j,k) = \frac{D(i,j+1,k) - D(i,j-1,k)}{\Delta y} \tag{3}$$

$$G_z(i,j,k) = \frac{D(i,j,k+1) - D(i,j,k-1)}{\Delta z} \tag{4}$$

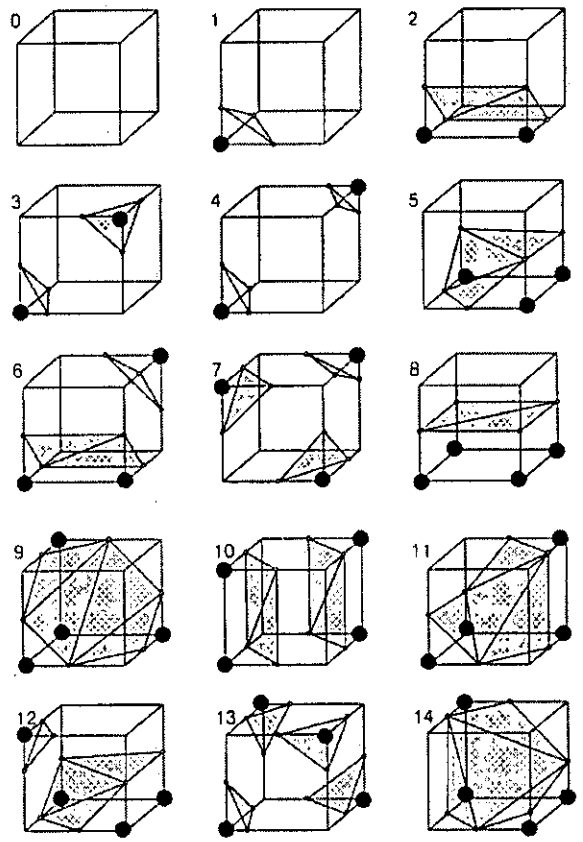


Figure 3. Triangulated Cubes.

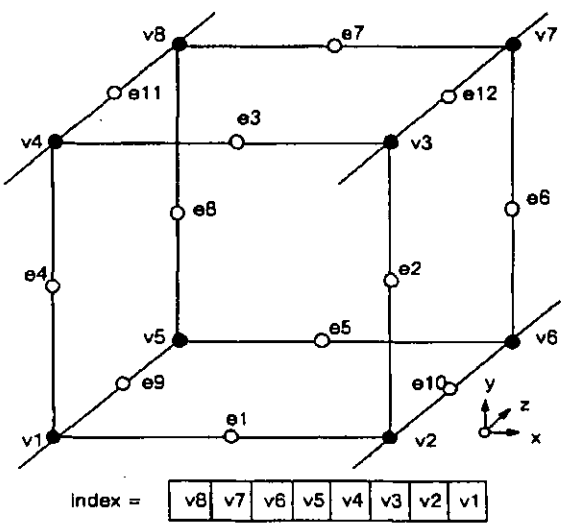


Figure 4. Cube Numbering.

where $D(i,j,k)$ is the density at pixel (i,j) in slice k and $\Delta x, \Delta y, \Delta z$ are the lengths of the cube edges. Dividing the gradient by its length produces the unit normal at the vertex required for rendering. We linearly interpolate this normal to the point of intersection. Note that to calculate the gradient at all vertices of the cube, we keep four slices in memory at once.

In summary, *marching cubes* creates a surface from a three-dimensional set of data as follows:

1. Read four slices into memory.
2. Scan two slices and create a cube from four neighbors on one slice and four neighbors on the next slice.
3. Calculate an index for the cube by comparing the eight density values at the cube vertices with the surface constant.
4. Using the index, look up the list of edges from a precalculated table.
5. Using the densities at each edge vertex, find the surface-edge intersection via linear interpolation.
6. Calculate a unit normal at each cube vertex using central differences. Interpolate the normal to each triangle vertex.
7. Output the triangle vertices and vertex normals.

5. ENHANCEMENTS TO THE BASIC ALGORITHM.

We have made several improvements to the original *marching cubes* that make the algorithm run faster and that add solid modeling capabilities.

5.1 Efficiency Enhancements.

The efficiency enhancements allow the algorithm to take advantage of pixel-to-pixel, line-to-line, and slice-to-slice coherence. For cubes interior to the original data limits (those not including slice 0, line 0, or pixel 0), only three new edges need to be interpolated for each cube. We can obtain the other nine edges from previous slices, lines, or pixels. In Figure 5, the shaded circles represent values available from prior calculations; only edges 6, 7, and 12 have to be calculated for the new cube.

Special cases are present along the boundaries of the data, but, by enumerating these cases, we can limit vertex calculations to once per vertex. In practice, we only save the previous pixel and line intersections because the memory required to save the previous slice's intersections is large. Using the coherence speeds up the algorithm by a factor of three.

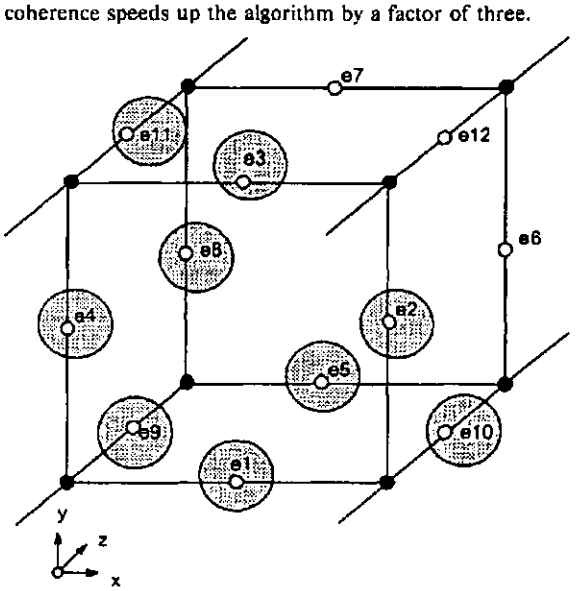


Figure 5. Coherence.

Reducing the slice resolution, by averaging four pixels into one, decreases the number of triangles, improves the surface construction efficiency and smooths the image. Although there is some loss of detail in the averaged slices, the averaging makes the number of triangles more manageable for high-resolution slices.

5.2 Functional Enhancements.

We have added a solid modeling capability to the algorithm. Boolean operations permit cutting and capping of solid models, as well as the extraction of multiple surfaces. In a medical application, cutting is analogous to performing surgery and capping (and texture mapping) is analogous to the medical imaging technique of reformatting.

We use the cube index described earlier to do Boolean operations on the surfaces. Here, just consider three values of the index:

index = *0* for cubes outside the surface.
index = *255* for cubes inside the surface.
0 < *index* < *255* for cubes on the surface.

Solid modeling uses these notions of *inside*, *outside*, and *on* to create a surface. Analytic functions also provide the same information; so, for example the equation of a plane, $ax + by + cz - d$, tells where a given point lies with respect to the plane. Let $-\mathbf{S}$, $\delta\mathbf{S}$, and $\mathbf{S}$ represent sets of points that are outside, on, and inside a surface, respectively. Referring to Figure 6, we build a truth table, shown in Figure 7, for the Boolean intersection operation.

Nine entries in the truth table describe what to do when two surfaces have a given index. With x's representing no operation, the entry for $(\mathbf{S}, -\mathbf{P})$ shows that the cube in question is inside one surface but outside the other, resulting in no triangles. The $(\delta\mathbf{S}, \mathbf{P})$ entry produces triangles from the *S* surface, while the $(\mathbf{S}, \delta\mathbf{P})$ entry produces triangles from the *P* surface. The $(\delta\mathbf{S}, \delta\mathbf{P})$ entry, created when a cube is on both surfaces, requires special processing. We clip

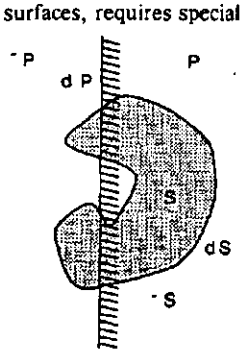


Figure 6. Point/Surface Relations.

	P	-P	d P
S	x	x	P
-S	x	x	x
dS	S	x	*

Figure 7. Truth Table.

each triangle from one surface against each triangle from the other, using the Sutherland-Hodgman clipping algorithm [28].

This technique applies to any surfaces that have inside/outside functions. We have used it with planes and with connectivity masks generated by separate image processing algorithms [9]. Application of a "logical or" truth table provides the capability for multiple surface extraction.

We implement texture mapping by finding the triangles on a plane's surface and attenuating the normal's length using the original slice data.

6. IMPLEMENTATION.

Marching cubes, written in C, runs on Sun Workstations[1] under Unix[2], VAX's under VMS[3], and an IBM 3081 under IX/370[4]. We display the models using an in-house z-buffer program or a General Electric Graphicon 700[5]. For our models, the Graphicon displays at a rate of 10,000 triangles per second. In addition to surfaces of constant density, the software allows any number of planes that can be transparent, capped with triangles, or textured with interpolated density data. Medical practitioners refer to this texture mapping as reformatting. Execution times depend on the number of surfaces and resolution of the original data. Model creation times on a VAX 11/780 vary from 100 seconds for 64 by 64 by 48 SPECT data to 30 minutes for 260 by 260 by 93 CT studies. Times for the same studies on the IBM 3081 are twelve times faster. The number of triangles in a surface model is proportional to the area of the surface. This number can get large (over 500,000 in some cases), so we reduce it using cut planes and surface connectivity. Also, sometimes we reduce the resolution of the original data by filtering, producing a somewhat smoother surface with some loss of resolution.

7. RESULTS.

We have applied *marching cubes* to data obtained from CT, MR, and SPECT, as well as data generated from analytic functions. We present three case studies that illustrate the quality of the constructed surfaces and some modeling options. Each image was rendered at 512 by 512 resolution without antialiasing.

7.1 Computed Tomography.

The first case is a CT study of the head of a twelve year old male with a hole in the skull near the left side of the nose. The 93 axial slices are 1.5 mm thick, with pixel dimensions of 0.8 mm. This study by D.C. Hemmy, MD, of the Medical College of Wisconsin, illustrates the detail present in surfaces constructed by *marching cubes*. Figures 8 and 9 show the bone and soft tissue surfaces respectively. The tube in the patient's mouth is present to administer anesthetic during the scanning process. The soft tissue image shows fine detail that includes the patient's pierced ear and the impression of adhesive tape on the face. Although these details are not clinically significant, they do show the resolution present in the constructed surface. Figure 10 is a tilted view of the soft tissue surface that shows nasal and ear passages. In Figure 11, a sagittal cut, texture mapped with the original

1 Sun Workstation is a trademark of Sun Microsystems.

2 Unix is a trademark of Bell Laboratories.

3 VAX and VMS are trademarks of Digital Equipment Corporation

4 IX/370 is a trademark of IBM.

5 Graphicon is a trademark of General Electric Company.

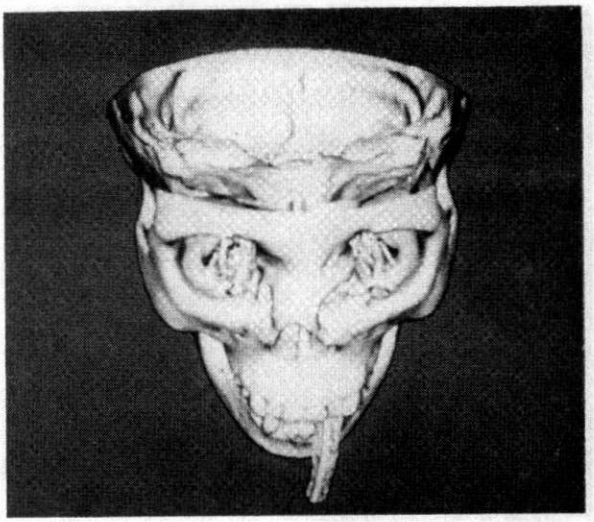


Figure 8. Bone Surface.

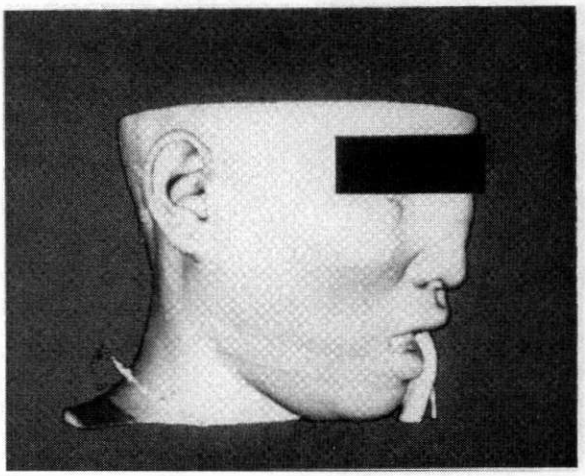


Figure 9. Soft Tissue Surface.

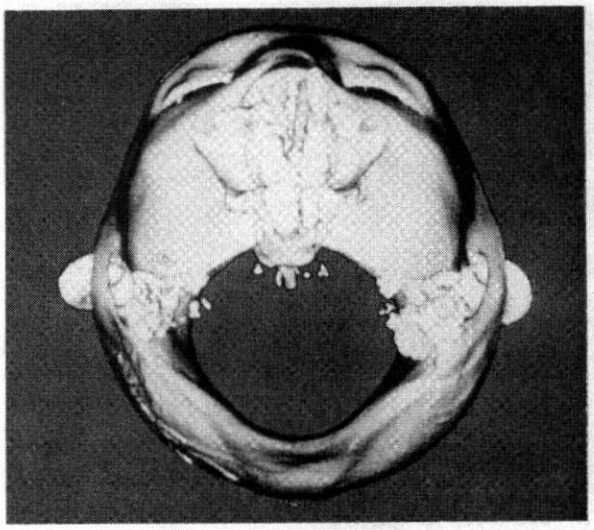


Figure 10. Soft Tissue, Top View.

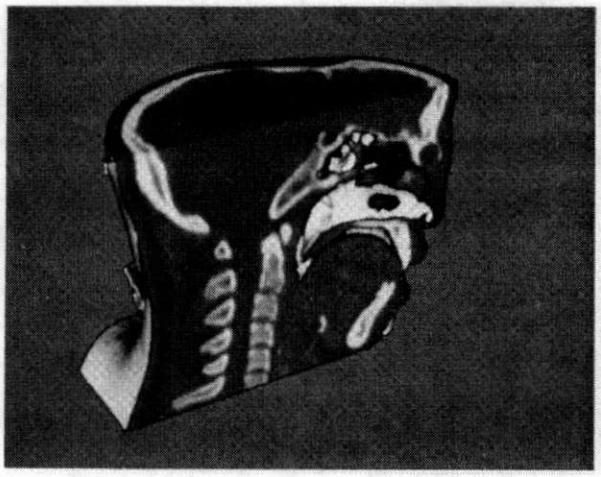


Figure 11. Sagittal Cut with Texture Mapping.

CT data, shows the slice data in relation to the constructed surface. The bone surface contains 550,000 triangles while the soft tissue surface has 375,000.

7.2 Magnetic Resonance.

The MR case of an adult male volunteer consists of 128 1.9 mm coronal slices. A 3D FT, flow compensated, fast sequence acquired the 128 slices in only 9 minutes. This pulse sequence, contrasting the unsaturated spins of the fresh blood flowing into the excited region of saturated spins, was produced by G. Glover of GE Medical Systems Group. Because of the complex anatomy present in the MR slices, we show, in Figure 12, the texture mapped cut surfaces intersected with the surface of the skin. Although the original slices are coronal, we show sagittal cuts to illustrate the algorithm's ability to interpolate texture on a cut plane. The largest surface model in the sequence contains 330,000 triangles, including triangles on the cut surface.

7.3 Single-Photon Emission Computed Tomography.

The SPECT study consisting of 29 coronal slices of the heart shows the algorithm's performance on low resolution data. D. Nowak from GE Medical Systems provided the 64 by 64 pixel data. Figure 13, showing the surface of the blood pool in the diastolic heart, contains 5,000 triangles. The descending aorta is the large vessel in the left of the picture.

8. CONCLUSIONS.

Marching cubes, a new algorithm for 3D surface construction, complements 2D CT, MR, and SPECT data by giving physicians 3D views of the anatomy. The algorithm uses a case table of edge intersections to describe how a surface cuts through each *cube* in a 3D data set. Additional realism is achieved by the calculation, from the original data, of the normalized gradient. The resulting polygonal structure can be displayed on conventional graphics display systems. Although these models often contain large numbers of triangles, surface cutting and connectivity can reduce this number. As CAD hardware increases in speed and capacity, we expect that *marching cubes* will receive increased use in practical, clinical environments.

Recently we developed another high-resolution surface construction algorithm called *dividing cubes* that generates points rather than triangles [10]. As the resolution of the 3D medical data increases, the number of triangles approaches the number of pixels in the displayed image. The density of surface points is chosen to cover the raster display. Both algorithms produce the same quality images, since the shading governs the perceived quality of the image.

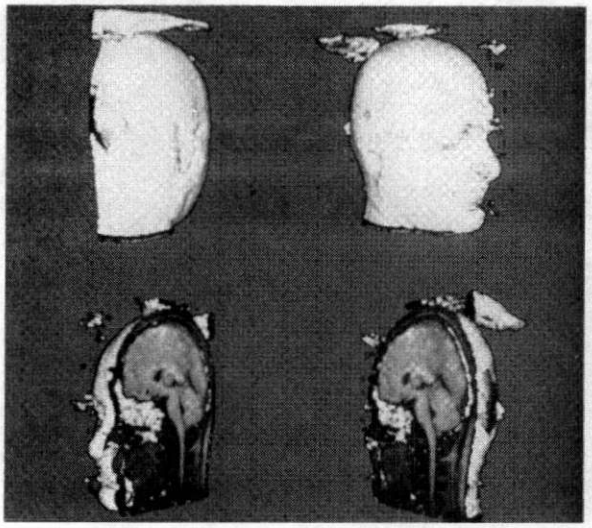


Figure 12. Rotated Sequence of Cut MR Brain.

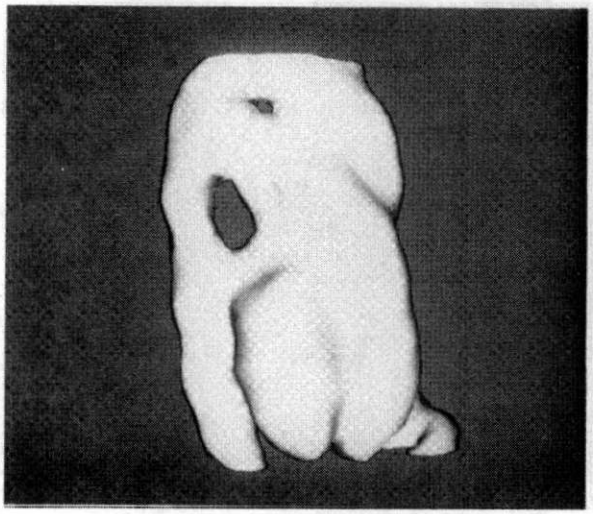


Figure 13. Blood Pool in the Diastolic Heart.

9. ACKNOWLEDGMENT.

We thank C. Crawford from General Electric's Medical Systems for stimulating our work in this area. Throughout the project, he has provided us with data and encouragement to improve the algorithm. R. Redington from our laboratory's Medical Diagnostics Branch provided a stable research environment and insight into the practical applications of 3D in medicine. W. Leue assisted us in converting between the different medical data formats and provided interfaces to our MR equipment.

10. REFERENCES

[1] Artzy, E., Frieder, G., and Herman, G. T. The Theory, Design, Implementation and Evaluation of a Three-Dimensional Surface Detection Algorithm. *Computer Graphics and Image Processing 15*, 1 (January 1981), 1-24.

[2] Barillot, C., Gibaud, B., Scarabin, J., and Coatrieux, J. 3D Reconstruction of Cerebral Blood Vessels. *IEEE Computer Graphics and Applications 5*, 12 (December 1985), 13-19.

[3] Bates, R. H., Garden, K. L., and Peters, T. M. Overview of Computerized Tomography with Emphasis on Future Developments. *Proc. of the IEEE* 71, 3 (March 1983), 356-372.

[4] Bloch, P. and Udupa, J. K. Application of Computerized Tomography to Radiation Therapy and Surgical Planning. *Proc. of the IEEE 71*, 3 (March 1983), 351-355.

[5] Brewster, L. J., Trivedi, S. S., Tut, H. K., and Udupa, J. K. Interactive Surgical Planning. *IEEE Computer Graphics and Applications 4*, 3 (March 1984), 31-40.

[6] Burk, D. L., Mears, D. C., Kennedy, W. H., Cooperstein, L. A., and Herbert, D. L. Three-Dimensional Computed Tomography of Acetabula Fractures. *Radiology 155*, 1 (1985), 183-186.

[7] Chen, L., Herman, G. T., Reynolds, R. A., and Udupa, J. K. Surface Shading in the Cuberille Environment. *IEEE Computer Graphics and Applications 5*, 12 (December 1985), 33-43.

[8] Christiansen, H. N. and Sederberg, T. W. Conversion of Complex Contour Line Definitions into Polygonal Element Meshes. *Computer Graphics 12*, 3 (August 1978), 187-192.

[9] Cline, H. E., Dumoulin, C. L., Lorensen, W. E., Hart, H. R., and Ludke, S. 3D Reconstruction of the Brain from Magnetic Resonance Images. *Magnetic Resonance Imaging* (1987, to appear).

[10] Cline, H. E., Lorensen, W. E., Ludke, S., Crawford, C. R., and Teeter, B. C. High-Resolution Three-Dimensional Reconstruction of Tomograms. *Medical Physics* (1987, to appear).

[11] Cook, L. T., Dwyer, S. J., Batnitzky, S., and Lee, K. R. A Three-Dimensional Display System for Diagnostic Imaging Applications. *IEEE Computer Graphics and Applications 3*, 5 (August 1983), 13-19.

[12] Farrell, E. J. Color Display and Interactive Interpretation of Three-Dimensional Data. *IBM J. Res. Develop 27*, 4 (July 1983), 356-366.

[13] Farrell, E. J., Zappulla, R., and Yang, W. C. Color 3D Imaging of Normal and Pathologic Intracranial Structures. *IEEE Computer Graphics and Applications 4*, 9 (September 1984), 5-17.

[14] Fuchs, H., Kedem, Z. M., and Uselton, S. P. Optimal Surface Reconstruction from Planar Contours. *Comm. of the ACM 20*, 10 (October 1977), 693-702.

[15] Gordon, D. and Reynolds, R. A. Image Space Shading of 3-Dimensional Objects. *Computer Graphics and Image Processing 29*, 3 (March 1985), 361-376.

[16] Hale, J. D., Valk, P. E., and Watts, J. C. MR Imaging of Blood Vessels Using Three-Dimensional Reconstruction: Methodology. *Radiology 157*, 3 (December 1985), 727-733.

[17] Hemmy, D. C., David, D. J., and Herman, G. T. Three-Dimensional Reconstruction of Craniofacial Deformity Using Computed Tomography. *Neurosurgery 13*, 5 (November 1983), 534-541.

[18] Hemmy, D. C. and Tessier, P. L. CT of Dry Skulls with Craniofacial Deformities: Accuracy of Three-Dimensional Reconstruction. *Radiology 157*, 1 (October 1985), 113-116.

[19] Herman, G. T. and Udupa, J. K. Display of 3D Digital Images: Computational Foundations and Medical Applications. *IEEE Computer Graphics and Applications 3*, 5 (August 1983), 39-46.

[20] Hinshaw, W. S. and Lent, A. H. An Introduction to NMR Imaging: From the Bloch Equation to the Imaging Equation. *Proc. of the IEEE 71*, 3 (March 1983), 338-350.

[21] Hoffman, E. A. and Ritman, E. L. Shape and Dimensions of Cardiac Chambers: Importance of CT Section Thickness and Orientation. *Radiology* 155, 3 (June 1985), 739-744.

[22] Hohne, K. H. and Bernstein, R. Shading 3D-Images from CT Using Gray-Level Gradients. *IEEE Trans. on Medical Imaging MI-5*, 1 (March 1986), 45-47.

[23] Keppel, E. Approximating Complex Surfaces by Triangulation of Contour Lines. *IBM J. Res. Develop 19*, 1 (January 1975), 2-11.

[24] Knoll, G. F. Single-Photon Emission Computed Tomography. *Proc. of the IEEE 71*, 3 (March 1983), 320-329.

[25] Meagher, D. J. Geometric Modeling Using Octree Encoding. *Computer Graphics and Image Processing 19*, 2 (June 1982), 129-147.

[26] Robb, R. A., Hoffman, E. A., Sinak, L. J., Harris, L. D., and Ritman, E. L. High-Speed Three-Dimensional X-Ray Computed Tomography: The Dynamic Spatial Reconstructor. *Proc. of the IEEE 71*, 3 (March 1983), 308-319.

[27] Sunguroff, A. and Greenberg, D. Computer Generated Images for Medical Application. *Computer Graphics 12*, 3 (August 1978), 196-202.

[28] Sutherland, I. E. and Hodgman, G. W. Reentrant Polygon Clipping. *Comm. of the ACM 17*, 1 (January 1974), 32-42.

[29] Trivedi, S. S., Herman, G. T., and Udupa, J. K. Segmentation Into Three Classes Using Gradients. *IEEE Trans. on Medical Imaging MI-5*, 2 (June 1986), 116-119.

[30] Udupa, J. K. Interactive Segmentation and Boundary Surface Formation for 3-D Digital Images. *Computer Graphics and Image Processing 18*, 3 (March 1982), 213-235.

[31] Vannier, M. W., Marsh, J. L., and Warren, J. O. Three Dimensional CT Reconstruction Images for Craniofacial Surgical Planning and Evaluation. *Radiology 150*, 1 (January 1984), 179-184.

[32] Zucker, S. W. and Hummel, R. A. A Three-Dimensional Edge Operator. *IEEE Trans. on Pattern Analysis and Machine Intelligence PAMI-3*, 3 (May 1981), 324-331.

IOP PUBLISHING PHYSICS IN MEDICINE AND BIOLOGY

Phys. Med. Biol. 53 (2008) 3175–3188 doi:10.1088/0031-9155/53/12/007

Comparative analysis of ^{60}Co intensity-modulated radiation therapy*

Christopher Fox[1,2], H Edwin Romeijn[2,3], Bart Lynch[2], Chunhua Men[3], Dionne M Aleman[4,6] and James F Dempsey[2,5,7]

[1] Sun Nuclear Corporation, 425-A Pineda Court, Melbourne, FL 32940, USA
[2] Department of Radiation Oncology, University of Florida, Gainesville, FL 32610-0385, USA
[3] Department of Industrial and Systems Engineering, University of Florida, Gainesville, FL 32611-6595, USA
[4] Department of Mechanical and Industrial Engineering, University of Toronto, Toronto, Ontario, Canada
[5] ViewRay Incorporated, 101 SE 2nd Place, Suite 201D, Gainesville, FL 32601, USA

E-mail: chrisfox@sunnuclear.com, romeijn@ise.ufl.edu, lynchb@ufl.edu, chhmen@ufl.edu, aleman@mie.utoronto.edu and dempsey@ufl.edu

Received 29 January 2008, in final form 1 April 2008
Published 27 May 2008
Online at stacks.iop.org/PMB/53/3175

Abstract

In this study, we perform a scientific comparative analysis of using ^{60}Co beams in intensity-modulated radiation therapy (IMRT). In particular, we evaluate the treatment plan quality obtained with (i) 6 MV, 18 MV and ^{60}Co IMRT; (ii) different numbers of static multileaf collimator (MLC) delivered ^{60}Co beams and (iii) a helical tomotherapy ^{60}Co beam geometry. We employ a convex fluence map optimization (FMO) model, which allows for the comparison of plan quality between different beam energies and configurations for a given case. A total of 25 clinical patient cases that each contain volumetric CT studies, primary and secondary delineated targets, and contoured structures were studied: 5 head-and-neck (H&N), 5 prostate, 5 central nervous system (CNS), 5 breast and 5 lung cases. The DICOM plan data were anonymized and exported to the University of Florida optimized radiation therapy (UFORT) treatment planning system. The FMO problem was solved for each case for 5–71 equidistant beams as well as a helical geometry for H&N, prostate, CNS and lung cases, and for 3–7 equidistant beams in the upper hemisphere for breast cases, all with 6 MV, 18 MV and ^{60}Co dose models. In all cases, 95% of the target volumes received at least the prescribed dose with clinical sparing criteria for critical organs being met for all structures that were not wholly or

* This work was supported by the National Science Foundation under grant no. DMI-0457394 and the National Cancer Institute under grant no. R01 CA100636.
[6] The work of this author was supported by The Alliance for Graduate Education and the Professoriate and a Graduate Research Fellowship of the National Science Foundation.
[7] This author owns stock in and is Chief Science Officer of ViewRay Incorporated and as such may benefit financially as a result of the outcomes of work or research reported in this manuscript.

0031-9155/08/123175+14$30.00

partially contained within the target volume. Improvements in critical organ sparing were found with an increasing number of equidistant ^{60}Co beams, yet were marginal above 9 beams for H&N, prostate, CNS and lung. Breast cases produced similar plans for 3–7 beams. A helical ^{60}Co beam geometry achieved similar plan quality as static plans with 11 equidistant ^{60}Co beams. Furthermore, 18 MV plans were initially found not to provide the same target coverage as 6 MV and ^{60}Co plans; however, adjusting the trade-offs in the optimization model allowed equivalent target coverage for 18 MV. For plans with comparable target coverage, critical structure sparing was best achieved with 6 MV beams followed closely by ^{60}Co beams, with 18 MV beams requiring significantly increased dose to critical structures. In this paper, we report in detail on a representative set of results from these experiments. The results of the investigation demonstrate the potential for IMRT radiotherapy employing commercially available ^{60}Co sources and a double-focused MLC. Increasing the number of equidistant beams beyond 9 was not observed to significantly improve target coverage or critical organ sparing and static plans were found to produce comparable plans to those obtained using a helical tomotherapy treatment delivery when optimized using the same well-tuned convex FMO model. While previous studies have shown that 18 MV plans are equivalent to 6 MV for prostate IMRT, we found that the 18 MV beams actually required more fluence to provide similar quality target coverage.

1. Introduction

Cobalt teletherapy units and linear accelerator systems (linacs) were introduced nearly simultaneously in the early 1950s and emerged as rival technologies for external beam therapy. The first two ^{60}Co units were installed in Canada in 1951 (Litt 2000) and the first clinical megavoltage (MV) linac was installed in London in 1952 (Bernier *et al* 2004, Thwaites and Tuohy 2006), with the first patient treated with this machine in 1953. The deeply penetrating ionizing photon beams provided by these devices quickly became the mainstay of radiation therapy, allowing for the noninvasive treatment of deep-seated tumors. In particular, both the linac and the cobalt teletherapy unit offered improved skin sparing and penetration over the orthovoltage unit employed until that time. Initially, cobalt teletherapy became the most widespread form of external beam therapy. This was mainly due to the safety, reliability, precision and simplicity of these units, requiring little maintenance and technical expertise to operate, as compared to the technologically intensive linacs. By the late 1960s, there were approximately 1700 external beam devices in the world and approximately 90% of them were cobalt therapy units (Hogstrom and Almond 2006). In the 1970s, major advances were made in the production of electron beams using linacs. Before the discovery and development of intensity-modulated radiation therapy (IMRT), electron beams were demonstrated to provide superior methods for treating cancer of the breast and the head-and-neck (H&N) (Hogstrom and Almond 2006). This, combined with a lack of technical improvements for the cobalt unit (lack of multileaf collimator (or MLC) and digital readout, etc), gave the linac a clinical advantage over the cobalt unit for nearly three decades. By the late 1980s, over 90% of therapy units in the US were linacs, and in the 1990s cobalt therapy units essentially vanished in the US. During this time, great advances were made in beam delivery based on large-scale treatment plan optimization, allowing the clinical application of IMRT using MLCs. With this

advent of MLC-based IMRT, the advantage of combining photon and electron beams vanished as IMRT was performed using only photon beams and could provide excellent treatment plans for cancer of the breast and head-and-neck (Webb 2004). In fact, so did the advantage of high-energy MV photon beams, as it has been demonstrated that IMRT only requires low-energy photon beams to produce high quality treatment plans, even for a deep-seated prostate tumor in exceptionally large patients (Pirzkall *et al* 2002, Sun and Ma 2006)).

Although low-energy photon beams have demonstrated efficacy for IMRT, cobalt units have other technical issues when compared with the linac (Laughlin *et al* 1986, Suit 1986). Historically, cobalt therapy systems were noted to suffer from four significant limitations:

(i) cobalt beams were noted to have a beam edge (or penumbra) that was not as sharp as that of a photon beam that can be produced from a linac;
(ii) cobalt beams often created high surface doses which could result in skin reactions due to low-energy contamination electrons (scattered from the source and collimators by the photon beam) (Mora *et al* 1999);
(iii) the most powerful cobalt beams had a lower dose rate than linacs by approximately 60% (with a maximum output of ~250 cGy min^{-1}) and
(iv) cobalt beams are not as penetrating as higher (10–20) MV beams available on a linac.

These disadvantages, along with the ability of electron beams to produce curative treatment plans for cancers of the breast and head-and-neck (in the pre-IMRT era), resulted in the eventual preference for the linac in the 1970s and 1980s and the eventual demise of cobalt teletherapy as a leading technology in the developed world.

However, cobalt was not without its supporters. In 1986, Laughlin *et al* (1986) published a paper detailing the pros and cons of ^{60}Co and called for a continued push in the technical development of cobalt therapy. That call was reiterated in an editorial by (Suit 1986) in the same journal issue, but went unheeded until now. Recently, development on a novel magnetic resonance image-guided radiation therapy (MRIGRT) device has begun (Renaissance™ System 1000, ViewRay Incorporated, Gainesville, OH) by the authors of this manuscript. The device is designed to overcome all four of the above-mentioned limitations of cobalt therapy in the following manner:

(i) the penumbra of a linac with a MLC (which has been measured to be in the range of 4–7 mm depending on manufacturer (Huq *et al* 2002, Kanagaki *et al* 2007, Langen *et al* 2005)) is actually comparable to that of a cobalt unit when a double-focused MLC is employed with a commercially available 2 cm diameter cobalt source (T1000, MDS Nordion; see the results in this paper);
(ii) the magnetic field of the MRI eliminates contamination electrons and with it the possibility of high surface doses and related skin reactions (Jursinic and Mackie 1996);
(iii) by utilizing three radiotherapy heads the MRIGRT device provides a competitive dose rate that is higher than a standard linac 2 years after source install and
(iv) the penetration of a ^{60}Co photon beam is not important with IMRT, which is supported by the results of this manuscript.

Previous work (Laughlin *et al* 1986) has considered the characteristics and merits of beam energies in the range 1–45 MV for conformal photon beams (non-IMRT) and demonstrated that for 3- and 4-beam planning equivalent dose distributions can be achieved with 6, 10 and 18 MV beams and clinical advantages associated with higher energy beams improve little beyond 4 MV. In the case of IMRT, most studies have demonstrated that 6 and 18 MV beams produce equivalent quality treatment plans (Pirzkall *et al* 2002, Sun and Ma 2006, Weiss *et al* 2007), although Madani *et al* (2007) observe some differences depending on the dose

calculation method used. However, comparisons of ^{60}Co and linac treatment planning for IMRT have not been published in the medical physics literature.

Consideration of the IMRT treatment setup must include not only the optimum beam quality but also a decision as to the number of beams and their orientation about the patient. This choice is often left to the discretion of the treatment center and is based on prior knowledge and experience. However, the fundamental question of the optimal number of equidistant beams and beam orientation optimization is still under discussion in the literature, although several studies have shown a sharply declining marginal benefit of using more than 10 equidistant beams for high-energy photon beam IMRT (see, e.g., Bortfeld *et al* (1990), Das *et al* (2003) and Stien *et al* (1997)). Additionally, rotating fan beams have been introduced that utilize a helical treatment pattern by rotating the beam about a 360° gantry and translating the patient couch through it (Beavis 2004, Yang *et al* 1997). This effectively approximates the limit of a large number of coplanar fan beams about the gantry for beam delivery.

The goal of this work is to investigate the use of different numbers of static multileaf collimator (MLC) delivered beams as well as a helical tomotherapy beam geometry in ^{60}Co IMRT. Treatment plans with 5–71 equidistant 6 MV and ^{60}Co beams are compared in terms of target coverage and organ sparing for H&N, prostate, CNS and lung cases, and 3–7 equidistant beams restricted to the upper hemisphere for breast cases. In addition, a helical beam implementation is compared to static field plans for a ^{60}Co beam modality for H&N cases. Finally, 6 MV, 18 MV and ^{60}Co beams were compared for a prostate case.

2. Materials and methods

Five patient cases for each of five typical IMRT treatment sites (H&N, prostate, CNS, breast and lung) were used in our study. H&N and CNS cases contained two targets, referred to as PTV1 and PTV2. For prostate cases, PTV1 contained the prostate gland plus an 8 mm isotropic margin to allow for spatial uncertainties arising from setup errors and physical motions during treatment. Similarly, PTV2 contained the prostate gland and seminal vesicles and an 8 mm isotropic margin. For lung cases, margins were added according to the tumor location. A symmetric margin of 0.5 cm was added to upper lobe tumors while to tumors in the lower lobe a 0.5 cm transversal and 1 cm craniocaudal margin was applied (as in the work by Leter *et al* (2005)). Targets for breast cases were designated as in the work by Hong *et al* (1999), where PTV1 comprised the tumor bed and PTV2 the breast tissue plus nodes with a 1 cm margin in the posterior direction and 2 cm in the superior and inferior. The PTV2 target was extended to the body contour in the anterior direction.

Prescription doses of 73.8 and 54 Gy were assigned to PTV1 and PTV2, respectively, for H&N, prostate and CNS cases. For lung cases, PTV1 and PTV2 were prescribed 70 and 50 Gy, respectively, and in breast cases the prescription doses for PTV1 and PTV2 were 60 and 54 Gy, respectively. Target coverage of 95% volume receiving the prescription dose ($D_{95\%} \geqslant D_{R_x}$) with the maximum dose limited to $1.1 \times D_{R_x}$ was deemed acceptable. The tolerance doses applied to critical structures are shown in table 1 and are given as maximum tolerance dose allowed or constraints on the dose per volume fraction.

Anonymized DICOM volumetric CT data and delineated targets and structures for all 25 cases were imported into the University of Florida optimized radiation therapy (UFORT) treatment planning decision support system (TPDSS). This treatment planning system has been commissioned for use with 6 MV, 18 MV and ^{60}Co beamlet models from a Varian 2100C/D linac and a Theratronics 1000C cobalt unit (with a commercially available cobalt source (13,000 Ci 2 cm, MDS Nordion) at a 1 m isocenter with a double-focused MLC with its furthest side at 50 cm from the source). These models were fitted to published data and

Table 1. Tolerance criteria for critical structures.

Organ	Criterion	Organ	Criterion
Retina/eye	<45 Gy	Eye lens	<12 Gy
Optic nerve	<50 Gy	Optic chiasm	<55 Gy
Brain stem	<55 Gy	Spinal cord	<45 Gy
Parotid gland	<30 Gy at 50%	Submandibular gland	<30 Gy at 50%
Rectum	<60 Gy at 30%	Bladder	<60 Gy at 30%
Skin	<60 Gy	Mandible	<70 Gy
Individual lung (breast)	<20 Gy at 20%	Total lung (lung)	<40 Gy at 40%
Individual lung (breast)	Mean < 15 Gy	Total lung (lung)	Mean < 15 Gy

validated with radiochromic film of 1×1 cm^2 beamlets formed by the accelerator jaws for 6 and 18 MV and by a Cerrobend block for ^{60}Co. Data were measured using the methods described by Dempsey *et al* (1999), (2000). For the dose calculations, the CT and structure data were mapped to an isotropic grid. Dempsey *et al* (2005) used a Fourier analysis to determine the required resolution for 6 MV beams. In that paper, the 80–20% penumbra for the 6 MV beam was 2.25 mm and the required resolution for 1% accuracy was found to be 2.5 mm. In addition, we found that the 80–20% penumbra for ^{60}Co was 4.5 mm and the required resolution for 1% accuracy was 4–5 mm. Hence, in order to ensure that all relevant information in the dose distribution was captured we employed voxels of size $2.5 \times 2.5 \times 2.5$ mm^3 for 6 and 18 MV and $4 \times 4 \times 4$ mm^3 for ^{60}Co. The UFORT TPDSS uses a point-in-polygon technique to associate structures with voxels in the grid and a combination of stereographic projection and three-dimensional ray-tracing is used to determine density scaled depths (see Fox *et al* (2006)) to account for heterogeneities and establish those voxels intersected by each beamlet that traverses the target volume. Relative intensities at intersected voxels from each beamlet are determined, yielding dose deposition coefficients D_{ijs}, representing the dose deposited per unit intensity from beamlet i to voxel j contained within structure s. These coefficients are then input into a fluence map optimization (FMO) model. As is standard in treatment plan optimization models, the dose d_{js} deposited in a voxel j contained within structure s is the result of the superposition of the intensity-weighted beamlet dose depositions, formally given by the linear expression

$$d_{js} = \sum_i D_{ijs} x_i$$

where x_i is the (nonnegative) intensity or weighting factor associated with beamlet i. Our FMO model (see, e.g., Romeijn *et al* (2003), (2004) and Tsien *et al* (2003)) is an analytic nonlinear convex model that employs voxel-based penalty functions. In particular, the objective function is formulated as

$$f = \sum_s \sum_j \left(\alpha_s^+ \max\left\{d_{js} - t_s^+, 0\right\}^{n_s^+} + \alpha_s^- \max\{t_s^- - d_{js}, 0\}^{n_s^-}\right)$$

where, for each structure s, α_s^+ and α_s^- are the importance weights assigned to overdosing and underdosing, t_s^+ and t_s^- are the thresholds associated with overdosing and underdosing, and n_s^+ and n_s^- are powers that ensure that the objective function value f is adversely and disproportionally affected as the dose moves away from the threshold values. Typically, both overdosing and underdosing penalties are included for targets while only overdosing penalties are included for critical structures. Targets are assigned the highest importance in the FMO model to ensure that the target coverage criteria are met, which is particularly important when overlap of targets and critical structures occurs. Similar tuning parameters were used for each

 30227

case with minor adjustments to improve target coverage and structure sparing as required. The optimal intensities of the available beamlets are then determined by optimizing the problem to provide a fluence map that gives a high-quality dose distribution satisfying the desired clinical criteria as described below.

FMO problems were solved to obtain treatment plans for H&N, prostate, CNS and lung cases using 5, 7, 9, 11, 17, 35 and 71 equidistant beam angles. Breast cases were restricted to a 180° arc above the breast to remove beamlets that enter the posterior to reach the tumor, and FMO problems were solved for 3, 5 and 7 equidistant beams. Plan quality was assessed via dose–volume histogram (DVH) and spatial dose distribution evaluation. A comparison of the plan quality between the 6 MV, 18 MV and ^{60}Co was performed and the impact of varying the number of equidistant coplanar beams on target coverage and critical structure sparing investigated. A helical scanning treatment beam was implemented in the UFORT TPDSS with the 6 MV and ^{60}Co beam models. A maximum collimator opening of 40×2 cm^2 was allowed with bixel size of 1×1 cm^2 and pitch 0.5. The start and end point of the helices was established from the limits of targets in the cranium–caudal orientation with an additional margin of 1 cm.

The convex FMO model was solved by employing a projected gradient algorithm with Armijo line search (Kelley 1987). This yields a fast method that produces excellent target coverage and organ sparing. Our implementation of the projected gradient algorithm proceeds as follows. In a given iteration, k, with current solution $x^{(k)}$, the next solution is given by

$$x^{(k+1)} = \max\{x^{(k)} - \lambda^* \nabla f_k, 0\}$$

where ∇f_k is the gradient of f evaluated at $x^{(k)}$ and λ^* is found by a search along the projected path $\max\{x^{(k)} - \lambda \nabla f_k, 0\}$ parameterized by the nonnegative step length λ. The search direction, being the negative of the gradient of f at the current solution, is a descent direction so that the result of the line search is a solution with a lower (i.e., improved) objective function value (see figure 1). The performance of the algorithm depends greatly on the step length. If the step length is too short the convergence of the algorithm will require a large number of iterations and will therefore be slow. On the other hand, if the step size is too large the optimal solution may be overshot, producing an insufficient reduction in the objective function in each step, again preventing fast convergence. In each iteration, a good step size can be obtained by (i) trying several test candidate values for λ and (ii) choosing the one that provides the largest improvement in objective function value. The Armijo line search method determines the step size by iteratively stepping backwards from a maximum step size along the search direction until a solution is found that attains a reduction in objective function value. Finally, the leaf-sequencing algorithm of Kamath *et al* (2003) was applied to the fluence maps obtained by the FMO to determine the number of MLC-shaped apertures required to successfully deliver the fluence map.

3. Results and discussion

3.1. ^{60}Co beamlet analysis

Figure 2 shows a profile of the radiochromic film measurement of a 2×2 cm^2 ^{60}Co beamlet measured on a Theratronics 1000C Cobalt unit at 0.5 cm depth in a 30×30 cm^2 solid water phantom at 100 cm source to surface distance using a divergent 10 cm thick Cerrobend collimator with its distal end at 50 cm from the source. The beamlet was measured on the central axis; however, the cobalt source has a very even energy spectrum and no flattening filter so that the off-axis beamlets are very similar. An 80–20% penumbra distance of

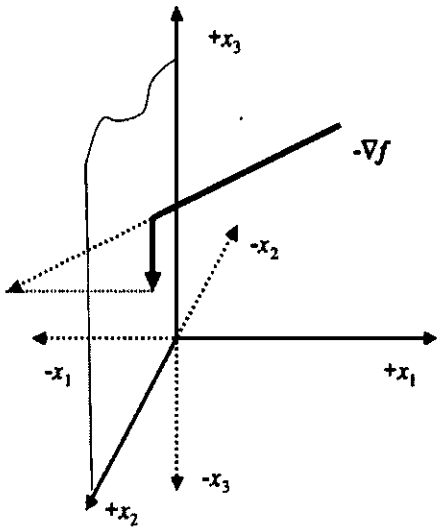


Figure 1. Example of a projected gradient step.

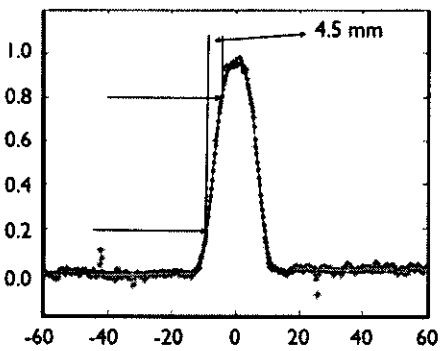


Figure 2. A 2 × 2 cm^2 ^{60}Co beamlet measured on a Theratronics 1000 C Cobalt unit.

4.5 mm was observed. Note that our results are consistent with the measured penumbra of cobalt sources reported in Adams and Warrington (2008), Poffenbarger and Podgorsak (1998), Raja Singh *et al* (2000) and Warrington and Adams (2002).

3.2. Comparative analysis of 6 MV and ^{60}Co beam IMRT plans

Figures 3(a), (c), (e), (g) and (i) show the DVHs for two 7-beam IMRT plans obtained for a typical case from each of the five sites studied (H&N, prostate, CNS, lung and breast), where the other cases showed comparable behavior. The solid lines represent plans for a 6 MV dose model while the dashed lines represent plans for a ^{60}Co dose model.

For the H&N case (figure 3(a)), with both 6 MV and ^{60}Co a treatment plan that achieves the required dose coverage of $D_{95\%} > D_{R_x}$ for the PTV1 and PTV2 was obtained while maintaining the submandibular and parotid glands within the criteria set in table 1. Target coverage is almost identical for both beam qualities. The four glands at the tolerance limits show minor variations between the two beams. The 6 MV beams produced slightly better

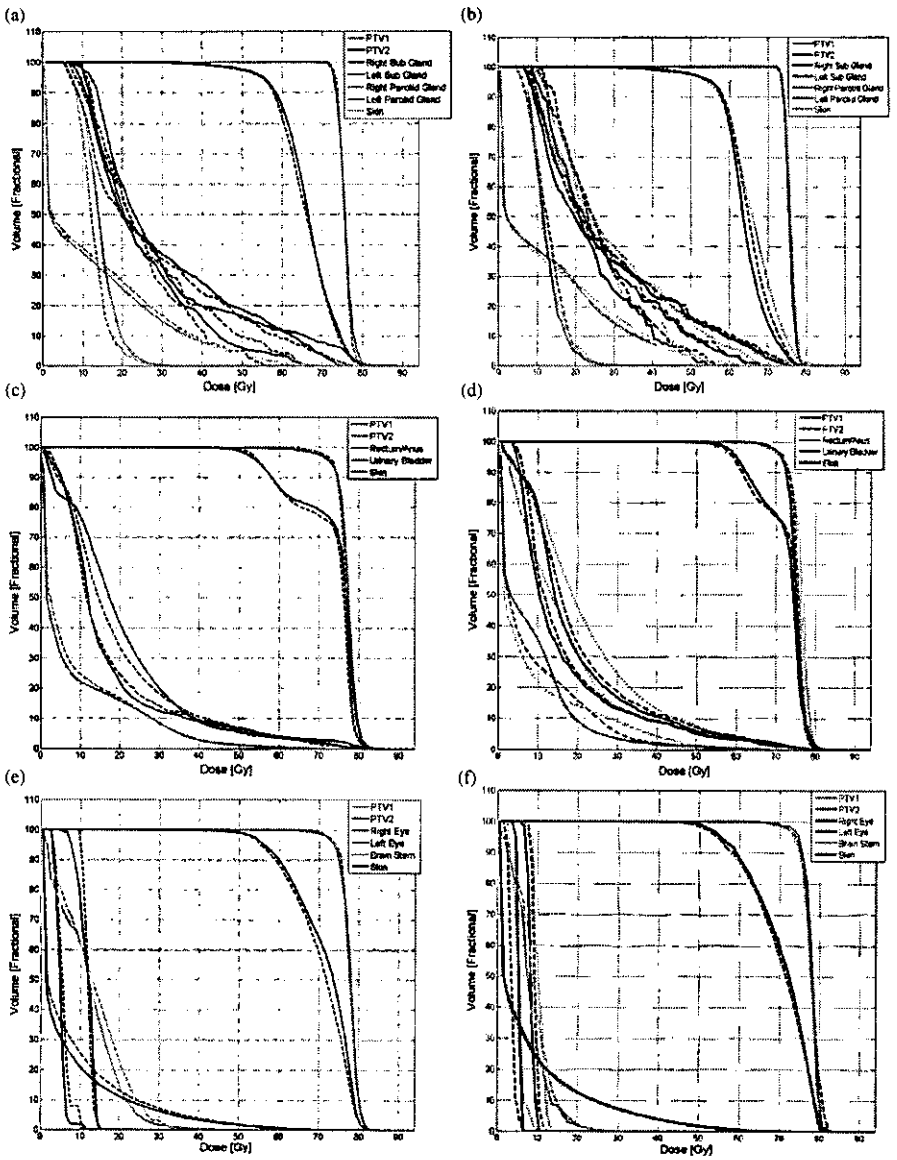


Figure 3. DVHs for typical clinical cases. Parts (a), (c) and (e) show a comparison of 6 MV (solid) and ^{60}Co (dashed) for 7 equidistant beams for H&N, prostate and CNS cases. Parts (b), (d) and (f) compare 5 (dotted), 9 (dashed) and 71 (solid) equidistant beams for the ^{60}Co dose model for H&N, prostate and CNS cases. Parts (g) and (i) show a comparison of 6 MV (solid) and ^{60}Co (dashed) for 7 equidistant beams for a lung case and 7 equidistant beams in the upper hemisphere for a breast case. Part (h) compares 5 (dotted), 9 (dashed) and 71 (solid) equidistant beams for the ^{60}Co dose model for a lung case, and part (j) compares 3 (dashed) and 5 (solid) equidistant beams in the upper hemisphere for the ^{60}Co dose model for a breast case.

sparing for left parotid and submandibular glands, at the sparing criterion condition, while ^{60}Co showed improved results for the right submandibular gland. These differences were not

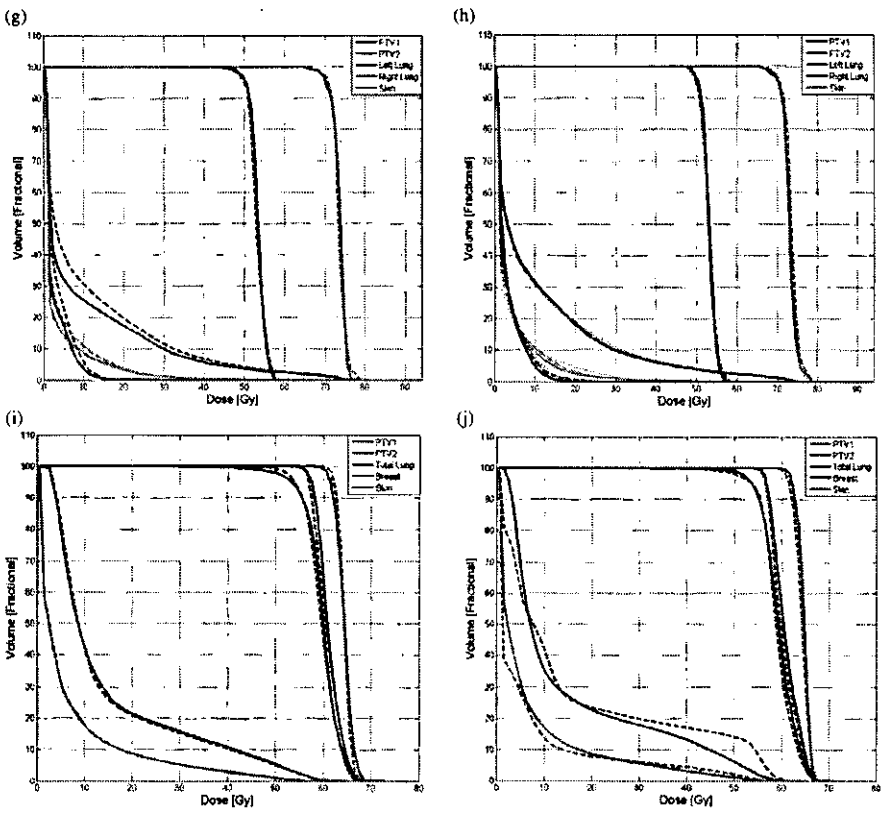


Figure 3. (Continued.)

deemed clinically significant and can be attributed to differences in the trade-offs made by the FMO model. The DVH above 30 Gy is observed to decrease more rapidly for ^{60}Co. At 50 Gy, the DVH for ^{60}Co is 3–4% lower compared to 6 MV and is most likely the result of the FMO models ability to take advantage of the reduced exit dose associated with the lower penetrability of the ^{60}Co beam.

For the prostate case (figure 3(c)), equivalent target coverage was observed with both 6 MV and ^{60}Co for 7-beam plans. Below 30 Gy, the rectum is observed to receive ~2–3 Gy less per volume than with 6 MV with the urinary bladder receiving approximately the same dose for each beam quality below 30 Gy. At the tolerance limits (<30% at 60 Gy), the dose from both modalities is the same for the urinary bladder and rectum. The lower maximum dose observed in the bladder and rectum using ^{60}Co results from the higher attenuation of the beam compared to 6 MV.

CNS cases are the simplest to solve, primarily due to the lack of proximity of the targets to critical structures. Figure 3(e) shows that the required target coverage of $D_{95\%} > D_{R_x}$ is achieved for both cases with only minor variations for 7-beam plans.

Figures 3(g) and (i) show a similar pattern for the lung and breast cases, respectively. In addition, for the breast case the mean dose to total lung was virtually identical at about 11 Gy for both 6 MV and ^{60}Co, while for the lung case marginally lower mean dose to both lungs was observed for 6 MV than for ^{60}Co (about 10 versus 11 Gy for right lung and about 2 versus

2.5 Gy for left lung), but all observed mean lung doses were well below the tolerance limit. In summary, in both cases plan quality is similar between the 6 MV and ^{60}Co modalities with 7 beams.

3.3. Comparative analysis of varying the number of beams

Figures 3(b), (d), (f), (h) and (j) show the DVHs for IMRT plans obtained for a typical case from each of the five sites with different numbers of equidistant ^{60}Co beams; the other cases exhibited similar behavior. In particular, in figures 3(b), (d), (f) and (h) (H&N, prostate, CNS and lung, respectively) the dotted lines correspond to 5-beam plans, the dashed lines to 9-beam plans and the solid lines to 71-beam plans, while in figure 3(j) (breast) the dashed lines correspond to a 3-beam plan while the solid lines correspond to a 5-beam plan.

For the H&N case (figure 3(b)), good target coverage is observed with as few as five equidistant coplanar beams while the sparing criteria given in table 1 are maintained. The threshold of 95% target volume receiving the prescription dose is attained for the PTV1 and PTV2 with 5, 9 and 71 beams. The PTV1 coverage is the same for each set of equidistant beams. However, the hot spot of the PTV2 at 20% volume improves by ~5% between the 5- and 9-beam plans. Further increasing the number of beams shows only minor improvements in the hot spot of the PTV2. Gland coverage shows a similar trend at 30 Gy with a reduction in the dose per volume of 3–10% when the number of beams is increased from 5 to 9 beams. Increasing the number of beams up to 71 shows little change as compared to 9 beams. Figure 4 illustrates the dose distributions obtained for 7 and 71 beams from the 6 MV and ^{60}Co dose models. Only minor variations are observed between the two treatment modalities, and similar dose distributions for targets and critical structures were observed for 7 and 71 beams.

As for the H&N case, the prostate plans demonstrate similar coverage of the PTV1 and PTV2 using 5, 9 and 71 beams (figure 3(d)). The dose to the urinary bladder and rectum is below accepted tolerances of 30% volume to receive less than 60 Gy for all three beam numbers. Below 40 Gy a marked increase in dose to critical structures is observed when the number of beams is decreased. The 5-beam plan results in ~5 Gy higher dose to 30% of the rectum volume as compared to the 9- and 71-beam plans, for which the corresponding dose is 20 Gy. The dose distribution of the skin varies considerably with changing number of beams. The 71-beam plan gives a lower dose to a larger percentage of the skin than the 5-beam plan: with the former plan 35% receives in excess of 10 Gy and about 3% in excess of 30 Gy, while with the latter plan only 20% of the skin receives in excess of 10 Gy and about 9% in excess of 30 Gy. This is likely the result of spreading the incident dose over a larger surface area with more beams. Figure 5 illustrates the dose distributions obtained for 7 and 71 beams from the 6 MV and ^{60}Co dose models. As for the H&N case, the variations observed between both the treatment modalities are small and clinically comparable dose distributions for targets and critical structures are obtained with 7 and 71 beams.

For both the CNS and the lung cases (figures 3(f) and (h)) only marginal improvements are observed when the number of beams increases from 5 to 9. For the lung case, the mean dose to either lung is virtually independent of the number of beams at 10–11 Gy for the right lung and <3 Gy for the left lung. Overall, all criteria are well within the tolerances and little further improvements are seen in the 71-beam plans over the 9-beam plans.

Finally, figure 3(j) shows the DVHs for 3-beam and a 5-beam ^{60}Co breast plan. (Note that beams for breast cases are restricted to a 180° arc above the breast tissue.) Comparing these DVHs with figure 3(i), which shows a 7-beam plan, the total lung dose is observed to fall off faster as the number of beams increases. In each case, the maximum dose received to

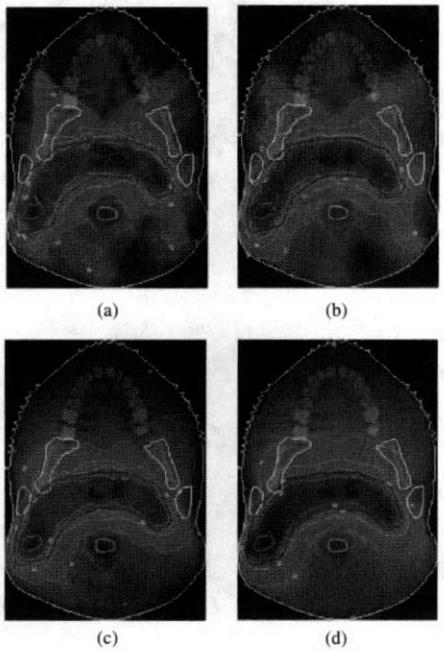


Figure 4. Deliverable IMRT axial isodose distributions for a H&N IMRT case. Parts (a) and (b) show 7-beam dose distributions for 6 MV and ^{60}Co, respectively. Parts (c) and (d) are 71-beam dose distributions for 6 MV and ^{60}Co, respectively. Shown are PTV1 (red) and PTV2 (dark blue) as well as the parotid glands (yellow), mandible (light green), spinal cord (light blue) and skin (bright green). Isodose curves are shown for 73.8, 60, 54, 45, 30 and 10 Gy.

the total lung is 60 Gy, the mean dose to lung is below the threshold at about 11–13 Gy, and the tolerance limit of no more than 20% lung volume receiving in excess of 40 Gy is also met.

3.4. Comparative analysis of 6 MV, 18 MV and ^{60}Co IMRT for prostate

Figure 6(a) shows an 18 MV dose model applied to a prostate case compared to 6 MV and ^{60}Co plans for the same case. The 18 MV beam model plan achieves the sparing criteria given in table 1 for the urinary bladder and rectum and the coverage of PTV1 and PTV2 attains the 95% volume receiving the prescription dose. However, the falloff in both the PTV2 and PTV1 is less sharp with the 18 MV beam and additional tuning of the objective function parameters produced little improvement and showed a marked increase of the rectum and urinary dose per volume around the 40% mark. The urinary bladder receives a consistently higher dose per volume with the 18 MV model, reaching a maximum of 20% volume having 20 Gy above that observed with the 6 MV case. This was also the case for the skin and parts of the rectum.

3.5. Comparative analysis of helical and static ^{60}Co for H&N

Figure 6(b) shows a H&N plan obtained using a helical ^{60}Co beam scanning pattern for the treatment delivery (dotted lines) as well as a plan for the same patient obtained with 11 static equidistant ^{60}Co beams (solid lines). Target coverage and critical structure sparing criteria are maintained for both plans. The target DVHs are almost identical, with the 11-beam static

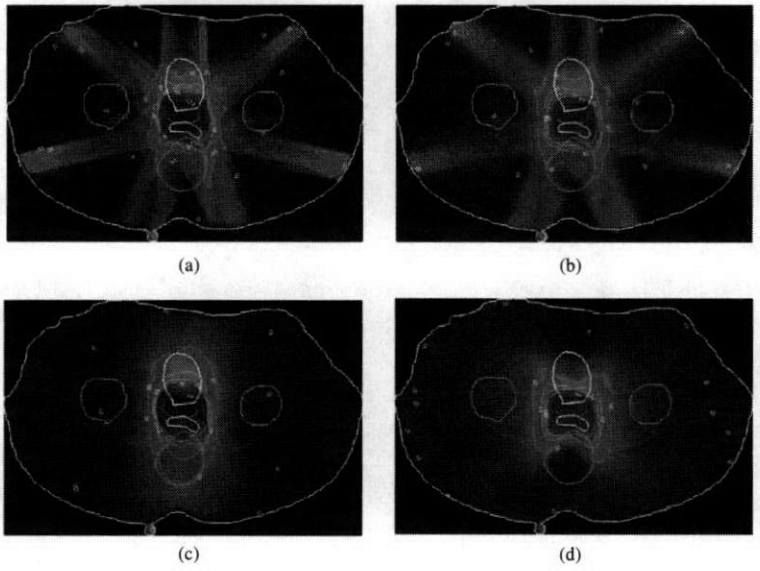


Figure 5. Deliverable IMRT axial isodose distributions for a prostate IMRT case. Parts (a) and (b) show 7-beam dose distributions for 6 MV and ^{60}Co, respectively. Parts (c) and (d) are 71-beam dose distributions for 6 MV and ^{60}Co, respectively. Shown are PTV1 (red) and PTV2 (dark blue) as well as bladder (yellow), rectum (brown), femoral heads (purple) and skin (bright green). Isodose curves are shown for 73.8, 60, 54, 45, 30 and 10 Gy.

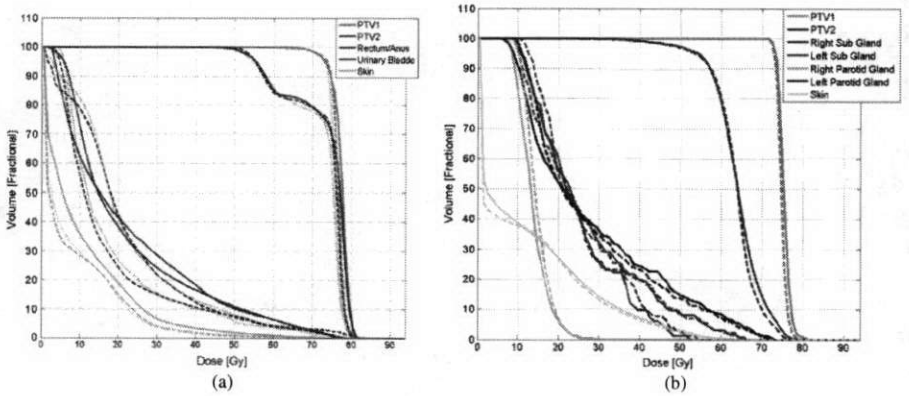


Figure 6. (a) Comparison of an 18 MV (solid), 6 MV (dashed) and ^{60}Co (dotted) dose model for a static 11-beam prostate plan. (b) Comparison of a static 11-beam ^{60}Co (solid line) plan to that of a helical ^{60}Co (dotted line) plan for a H&N case.

plan showing a maximum divergence at the 15% volume level of 2 Gy for PTV2 and 1 Gy for PTV1, while the critical structure DVHs are virtually indistinguishable from one another.

3.6. Number of apertures

Finally, the leaf-sequencing algorithm was applied to the 25 cases investigated. To illustrate these results, table 2 shows, for a single case for each of the five sites and for both 6 MV and

Table 2. Number of apertures required for typical cases.

	H&N		Prostate		CNS		Lung		Breast	
Number of beams	6 MV	^{60}Co	6 MV	^{60}Co	6 MV	^{60}Co	6 MV	^{60}Co	6 MV	^{60}Co
3									89	89
5	164	150	130	135	108	107	104	107	157	137
7	226	224	167	209	150	151	257	276	192	180
9	283	316	219	238	188	195	328	353		
11	354	374	282	283	233	237	409	432		
17	504	537	397	480	346	358	630	662		
35	1014	1178	890	965	788	770	1249	1385		
71	2330	2455	1743	1976	1623	1569	2561	2969		

^{60}Co beams, the number of apertures required to deliver the fluence map obtained from the FMO using the leaf-sequencing algorithm. The number of apertures varied from ~90 for a 3-beam breast plan up to ~3000 for a 71-beam lung plan. For typical 7-beam plans, the H&N cases required on the order of 225 apertures, compared to approximately 200 for prostate, 150 for CNS and 275 for lung. In general, the number of apertures required increased linearly with the number of beams for all sites and for both beam qualities. No trend was observed between the target volume and aperture number. Finally, fluence maps that contain a larger number of gradient changes across the profile generally require a larger number of apertures to successfully deliver the FMO fluence map using a MLC.

4. Conclusions

The data presented demonstrates that excellent plan quality for IMRT using inverse treatment planning can be achieved with low numbers of equidistant beams, with little gain from extending beam numbers beyond 9 beams in terms of target coverage and critical organ sparing. We also demonstrate the feasibility of employing commercial ^{60}Co sources with a divergent MLC for IMRT and show that nearly identical plans can be achieved when compared to 6 MV IMRT. We therefore conclude that the common assumption that the ^{60}Co penumbra is inferior to linac penumbra for MLC-based IMRT is not supported by the literature (see, e.g., Huq *et al* (2002), Kanagaki *et al* (2007), Langen *et al* (2005) or the findings of this study.

References

Adams E J and Warrington A P 2008 A comparison between cobalt and linear accelerator-based treatment plans for conformal and intensity-modulated radiotherapy *Br. J. Radiol.* **81** 304–10

Beavis A W 2004 Is tomotherapy the future of IMRT? *Br. J. Radiol.* **77** 285–95

Bernier J, Hall E J and Giaccia A 2004 Radiation oncology: a century of achievements *Nature Rev. Cancer* **4** 737–47

Bortfeld T, Burkelbach J, Boesecke R and Schlegel W 1990 Methods of image reconstruction from projections applied to conformal radiotherapy *Phys. Med. Biol.* **25** 435–43

Das S, Cullip T, Tracton G, Chang A, Marks L, Anscher M and Rosenman J 2003 Beam orientation selection for intensity-modulated radiation therapy based on target equivalent uniform dose maximization *Int. J. Radiat. Oncol. Biol. Phys.* **55** 215–24

Dempsey J F, Low D A, Kirov A S and Williamson J F 1999 Quantitative optical densitometry with scanning-laser film digitizers *Med. Phys.* **26** 1721–31

Dempsey J F, Low D A, Mutic S, Markman J, Kirov A S, Nussbaum G H and Williamson J F 2000 Validation of a precision radiochromic film dosimetry system for quantitative two-dimensional imaging of acute exposure dose distributions *Med. Phys.* **27** 2462–75

Dempsey J F, Romeijn H E, Li J G, Low D A and Palta J R 2005 A Fourier analysis of the dose grid resolution required for accurate IMRT fluence map optimization *Med. Phys.* **32** 380–8

Fox C, Romeijn H E and Dempsey J F 2006 Fast voxel and polygon ray-tracing algorithms in IMRT treatment planning *Med. Phys.* **33** 1364–71

Hogstrom K R and Almond P R 2006 Review of electron beam therapy physics *Phys. Med. Biol.* **51** R455–89

Hong L, Hunt M, Chui C, Spirou S, Forester K, Lee H, Yahalom J, Kutcher G J and McCormick B 1999 Intensity-modulated tangential beam irradiation of the intact breast *Int. J. Radiat. Oncol. Biol. Phys.* **44** 1155–64

Huq M S, Das I J, Steinberg T and Galvin J M 2002 A dosimetric comparison of various multileaf collimators *Phys. Med. Biol.* **47** N159–70

Jursinic P A and Mackie T R 1996 Characteristics of secondary electrons produced by 6, 10 and 24 MV x-ray beams *Phys. Med. Biol.* **41** 1499–509

Kamath S, Sahni S, Li J, Palta J and Ranka S 2003 Leaf sequencing algorithms for segmented multileaf collimation *Phys. Med. Biol.* **48** 307–24

Kanagaki B, Read P W, Molloy J A, Larner J M and Sheng K 2007 A motion phantom study on helical tomotherapy: the dosimetric impacts of delivery technique and motion *Phys. Med. Biol.* **52** 243–55

Kelley C T 1987 *Iterative Methods in Optimization* (Philadelphia: SIAM)

Langen K M, Meeks S L, Poole D O, Wagner T H, Willoughby T R, Zeidan O A, Kupelian P A, Ruchala K J and Oliveira G H 2005 Evaluation of a diode array for QA measurements on a helical tomotherapy unit *Med. Phys.* **32** 3424–30

Laughlin J S, Mohan R and Kutcher G J 1986 Choice of optimum megavoltage for accelerators for photon beam treatment *Int. J. Radiat. Oncol. Biol. Phys.* **12** 1551–7

Leter E M, Cademartiri F, Levendag P C, Flohr T, Stam H and Nowak P J 2005 Four-dimensional multi slice computed tomography for determination of respiratory lung tumor motion in conformal radiotherapy *Int. J. Radiat. Oncol. Biol. Phys.* **62** 888–92

Litt P 2000 *Isotopes and Innovation: MDS Nordion's First Fifty Years, 1946–1996* (Montreal: McGill-Queens University Press)

Madani I, Vanderstraeten B, Bral S, Coghe M, De Gersem W, De Wagter C, Thierens H and De Neve W 2007 Comparison of 6 MV and 18 MV photons for IMRT treatment of lung cancer *Radiother. Oncol.* **82** 63–9

Mora G M, Maio A and Rogers D W O 1999 Monte Carlo simulation of a typical 60 Co therapy source *Med. Phys.* **26** 2494–502

Pirzkall A, Carol M P, Pickett B, Xia P, Roach M III and Verhey L J 2002 The effect of beam energy and number of fields on photon-based IMRT for deep-seated targets *Int. J. Radiat. Oncol. Biol. Phys.* **53** 434–42

Poffenbarger B A and Podgorsak E B 1998 Viability of an isocentric cobalt-60 teletherapy unit for stereotactic radiosurgery *Med. Phys.* **25** 1935–43

Raja Singh R, Ravindran P, Nizin P S and Ayyangar K 2000 Dosimetric study of the narrow beams of ^{60}Co teletherapy unit for stereotactic radiosurgery *Med. Dosim.* **25** 163–9

Romeijn H E, Ahuja R K, Dempsey J F, Kumar A and Li J G 2003 A novel linear programming approach to fluence map optimization for intensity modulated radiation therapy treatment planning *Phys. Med. Biol.* **48** 3521–42

Romeijn H E, Dempsey J F and Li J G 2004 A unifying framework for multi-criteria fluence map optimization models *Phys. Med. Biol.* **49** 1991–2013

Stien J, Mohan R, Wang X-H, Bortfeld T, Wu Q, Preiser K, Ling C C and Schlegel W 1997 Number and orientations of beams in intensity-modulated radiation treatments *Med. Phys.* **24** 149–60

Suit H 1986 What's the optimum choice? *Int. J. Radiat. Oncol. Biol. Phys.* **12** 1711–2

Sun M and Ma L 2006 Treatments of exceptionally large prostate cancer patients with low-energy intensity-modulated photons *J. Appl. Clin. Med. Phys.* **7** 43–9

Thwaites D I and Tuohy J B 2006 Back to the future: the history and development of the clinical linear accelerator *Phys. Med. Biol.* **51** R343–62

Tsien C, Eisbruch A, McShan D, Kessler M, Marsh R and Fraass B 2003 Intensity-modulated radiation therapy (IMRT) for locally advanced paranasal sinus tumors: incorporating clinical decisions in the optimization process *Int. J. Radiat. Oncol. Biol. Phys.* **55** 776–84

Warrington A P and Adams E J 2002 Cobalt 60 teletherapy for cancer: a revived treatment modality for the 21st century *IEE Seminar on Appropriate Medical Technology for Developing Countries*

Webb S 2004 *Contemporary IMRT: Developing Physics and Clinical Implementation* (Madison, WI: Advanced Medical Publishing)

Weiss E, Siebers J V and Keall P J 2007 An analysis of 6-MV versus 18-MV photon energy plans for intensity-modulated radiation therapy (IMRT) of lung cancer *Radiother. Oncol.* **82** 55–62

Yang J, Mackie T R, Reckwerdt P, Deasy J O and Thomadsen B R 1997 An investigation of tomotherapy beam delivery *Med. Phys.* **24** 425–36

[redacted]

GRAPHICS GEMS

edited by

ANDREW S. GLASSNER

Xerox Palo Alto Research Center
Palo Alto, California

AP PROFESSIONAL
AP Professional is a division of Academic Press, Inc.

Boston San Diego New York
London Sydney Tokyo Toronto

TRANSFORMING AXIS-ALIGNED BOUNDING BOXES

James Arvo
Apollo Systems Division
of Hewlett-Packard
Chelmsford, Massachusetts

A very common type of three-dimensional bounding volume is the axis-aligned box, a parallelepiped with each face perpendicular to one coordinate axis. The appeal of this shape is its simplicity. It is ubiquitous in ray tracing because it is among the simplest objects to test for ray intersection. It is also widely used to accelerate the rendering of display lists by facilitating quick visibility tests for collections of drawing primitives.

In both contexts it is frequently necessary to construct a bounding box of an object to which an affine transformation has been applied, typically by means of a 3×3 modeling matrix, M, followed by a translation, T. A simple and frequently acceptable means of constructing such a box is to transform the bounding box of the original object and enclose the resulting arbitrary parallelepiped by an axis-aligned box. This is equivalent to transforming the eight vertices of the original box and finding the extrema of the resulting coordinates. Since each point transformation requires nine multiplies and nine adds, this would entail 144 arithmetic operations and a minimum of 21 compares. This naive approach is wasteful because it ignores the information embodied in the cube's symmetry. We will show how to take advantage of this information.

We address two common methods of encoding a bounding box, B. The first is the use of three intervals, $[B_x^{min}, B_x^{max}]$, $[B_y^{min}, B_y^{max}]$, $[B_z^{min}, B_z^{max}]$. Aside from ordering, this is equivalent to storing two opposing vertices. The second method is to store the box center, $(B_x^{cent}, B_y^{cent}, B_z^{cent})$, and the box *half-diagonal*, $(B_x^{diag}, B_y^{diag}, B_z^{diag})$, which is the positive vector from the center of the box to the vertex with the three largest components. Both of these representations are amenable to very efficient transformation.

The algorithm shown in Fig. 1 transforms box A, encoded as intervals, into another axis-aligned box, B, of the same form. The algorithm is based on the following observation. To compute a component of the transformed box, say, the maximum along the ith axis, we need only consider which of the eight vertices produces the maximal product with the ith row of the matrix. There are two possibilities for each component of the potential vertex: the minimum or the maximum of the interval for that axis. By forming both products for each component and summing the largest terms, we arrive at the maximal value. The minimal value is found by summing the smaller terms. The translation component of the matrix does not influence these choices and is simply added in.

The algorithm shown in Fig. 2 transforms box A, now encoded as a center and half-diagonal vector, into another axis-aligned box, B, of the same form. In this form the new center, B^{cent}, is obtained by simply applying the affine transformation to A^{cent}. The ith component of the new half-diagonal, B_i^{diag}, is obtained by selecting the signed half-

```
procedure Transform_Interval_Box(M, T, A, B)
begin
    for i = 1...3 do

        Start with a degenerate interval at T_i to account for translation.

        B_i^min ← T_i;
        B_i^max ← T_i;

        Add in extreme values obtained by computing the products of the
        mins and maxes with the elements of the i'th row of M

        for j = 1...3 do
            a ← M_i,j * A_j^min;
            b ← M_i,j * A_j^max;
            B_i^min ← B_i^min + min(a, b);
            B_i^max ← B_i^max + max(a, b);
            endloop;
        endloop;
    end;
```

Figure 1. An algorithm for transforming an axis-aligned bounding box, A, stored as three intervals into another box, B, of the same form.

```
procedure Transform_CenterDiag_Box(M, T, A, B)
begin
    for i = 1...3 do

        Initialize the output variables by zeroing the new half-diagonal and
        setting the new center equal to the translation T.

        B_i^cent ← T_i;
        B_i^diag ← 0;

        Compute the i'th coordinate of the center by adding M_i,* · A^cent,
        and the i'th coordinate of the half-diagonal by adding |M_i,*| · A^diag.

        for j = 1...3 do
            B_i^cent ← B_i^cent + M_i,j * A_j^cent;
            B_i^diag ← B_i^diag + |M_i,j| * A_j^diag;
            endloop;
        endloop;
    end;
```

Figure 2. An algorithm for transforming an axis-aligned bounding box, A, stored as a center and a half-diagonal into another box, B, of the same form.

diagonal of A, which results in the maximal product with the ith row of M. Here "signed" means allowing each component to be either positive or negative independently. This generates all eight half-diagonals of box A, pointing from A^{cent} to each vertex. We achieve the maximum product with the row of M by making each of its three terms positive, negating the negative elements of M. Because A^{cent} is a positive vector, this is equivalent to taking the absolute value of each element of M, as shown in Fig. 2.

The cost of both of these algorithms is only 36 arithmetic operations and 9 compares. Note that in the first algorithm both $min(a, b)$ and $max(a, b)$ can be computed with one compare, and in the second algorithm each absolute value is counted as one compare.

See Appendix 2 for C Implementation (785)

Second Edition

Discrete Mathematics for Computer Scientists and Mathematicians

Joe L. Mott
Abraham Kandel
Theodore P. Baker
The Florida State University
Department of Mathematics and Computer Science

A Reston Book
Prentice-Hall
Englewood Cliffs, New Jersey 07632

It is important to realize that a set may itself be an element of some other set. For example, a line is a set of points; the set of all lines in the plane is a set of sets of points. In fact a set can be a set of sets of sets and so on. The theory dealing with the (abstract) sets defined in the above manner is called **(abstract or conventional) set theory,** in contrast to fuzzy set theory which will be introduced later in Chapter 8.

This chapter begins with a review of set theory which includes the introduction of several important classes of sets and their properties.

In this chapter we also introduce the basic concepts of relations and functions necessary for understanding the remainder of the material. The chapter also describes different methods of proof—including mathematical induction—and shows how to use these techniques in proving results related to the content of the text.

The material in Chapters 2–8 represents the applications of the concepts introduced in this chapter. Understanding these concepts and their potential applications is good preparation for most computer science and mathematics majors.

1.2 SETS AND OPERATIONS OF SETS

Sets will be denoted by *capital* letters $A,B,C,\ldots,X,Y,Z$. Elements will be denoted by *lower case* letters $a,b,c,\ldots,x,y,z$. The phrase "is an element of" will be denoted by the symbol $\in$. Thus we write $x \in A$ for "x is an element of A." In analogous situations, we write $x \notin A$ for "x is not an element of A."

There are five ways used to describe a set.

1. Describe a set by describing the properties of the members of the set.
2. Describe a set by listing its elements.
3. Describe a set A by its characteristic function, defined as

$$\mu_A(x) = 1 \text{ if } x \in A,$$

$$\mu_A(x) = 0 \text{ if } x \notin A,$$

for all x in U, where U is the universal set, sometimes called the "universe of discourse," or just the "universe," which is a fixed specified set describing the context for the duration of the discussion.

If the discussion refers to dogs only, for example, then the universe of discourse is the class of dogs. In elementary algebra or number theory,

the universe of discourse could be numbers (rational, real, complex, etc.). The universe of discourse must be explicitly stated, because the truth value of a statement depends upon it, as we shall see later.

4. Describe a set by a recursive formula. This is to give one or more elements of the set and a rule by which the rest of the elements of the set may be generated. We return to this idea in Section 1.10 and in Chapter 3.

5. Describe a set by an operation (such as union, intersection, complement, etc.) on some other sets.

Example 1.2.1. Describe the set containing all the nonnegative integers less than or equal to 5.

Let A denote the set. Then the set A can be described in the following ways:

1. $A = \{x \mid x \text{ is a nonnegative integer less than or equal to } 5\}$.
2. $A = \{0,1,2,3,4,5\}$.
3. $\mu_A(x) = \begin{cases} 1 \text{ for } x = 0,1,\ldots,5, \\ 0 \text{ otherwise.} \end{cases}$
4. $A = \{x_{i+1} = x_i + 1, i = 0,1,\ldots,4, \text{ where } x_0 = 0\}$.
5. This part is left to the reader as an exercise to be completed once the operations on sets are discussed.

The use of braces and | ("such that") is a conventional notation which reads: $\{x \mid \text{property of } x\}$ means "the set of all elements x *such that* x has the given property." Note that, for a given set, not all the five ways of describing it are always possible. For example, the set of real numbers between 0 and 1 cannot be described by either listing all its elements or by a recursive formula.

In this section, we shall introduce the fundamental operations on sets and the relations among these operations. We begin with the following definitions.

Definition 1.2.1. Let A and B be two sets. Then A is said to be a **subset** of B if every element of A is an element of B; A is said to be a proper subset of B if A is a subset of B and there is at least one element of B which is not in A.

If A is a subset of B, we say A is contained in B. Symbolically, we write $A \subseteq B$. If A is a proper subset of B, then we say A is strictly contained in

B, denoted by $A \subset B$. The containment of sets has the following properties. Let A, B, and C be sets.

1. $A \subseteq A$.
2. If $A \subseteq B$ and $B \subseteq C$, then $A \subseteq C$.
3. If $A \subseteq B$ and $B \subset C$, then $A \subset C$.
4. If $A \subseteq B$ and $A \not\subseteq C$, then $B \not\subseteq C$, where $\not\subseteq$ means "is not contained in."

The statement $A \subseteq B$ does not rule out the possibility that $B \subseteq A$. In fact, we have both $A \subseteq B$ and $B \subseteq A$ if and only if (abbreviated iff) A and B have the same elements. Thus we define the following:

Definition 1.2.2. Two sets A and B are equal iff $A \subseteq B$ and $B \subseteq A$. We write $A = B$.

Therefore, we have the following principle.

Principle. To show that two sets A and B are equal, we must show that each element of A is also an element of B, and conversely.

A set containing no elements is called the **empty set** or **null set**, denoted by $\varnothing$. For example, given the universal set U of all positive numbers, the set of all positive numbers x in U satisfying the equation $x + 1 = 0$ is an empty set since there are no positive numbers which can satisfy this equation. The empty set is a subset of every set. In other words, $\varnothing \subseteq A$ for every A. This is because there are no elements in $\varnothing$; therefore, every element in $\varnothing$ belongs to A. It is important to note that the sets $\varnothing$ and $\{\varnothing\}$ are very different sets. The former has no elements, whereas the latter has the unique element $\varnothing$. A set containing a single element is called a **singleton.**

We shall now describe three operations on sets; namely, complement, union, and intersection. These operations allow us to construct new sets from given sets. We shall also study the relationships among these operations.

Definition 1.2.3. Let U be the universal set and let A be any subset of U. The **absolute complement** of A, $\overline{A}$, is defined as $\{x \mid x \notin A\}$ or, $\{x \mid x \in U \text{ and } x \notin A\}$. If A and B are sets, the **relative complement** of A with respect to B is as shown below.

$$B - A = \{x \mid x \in B \text{ and } x \notin A\}.$$

It is clear that $\overline{\varnothing} = U$, $\overline{U} = \varnothing$, and that the complement of the complement of A is equal to A.

Definition 1.2.4. Let A and B be two sets. The **union** of A and B is $A \cup B = \{x \mid x \in A \text{ or } x \in B \text{ or both}\}$. More generally, if $A_1, A_2, \ldots, A_n$ are

sets, then their union is the set of all objects which belong to at least one of them, and is denoted by

$$A_1 \cup A_2 \cup \cdots \cup A_n, \text{ or by } \bigcup_{j=1}^{n} A_j.$$

Definition 1.2.5. The **intersection** of two sets A and B is $A \cap B = \{x \mid x \in A \text{ and } x \in B\}$. The intersection of n sets $A_1, A_2, \ldots, A_n$ is the set of all objects which belong to every one of them, and is denoted by

$$A_1 \cap A_2 \cap \cdots \cap A_n, \text{ or } \bigcap_{j=1}^{n} A_j.$$

Some basic properties of union and intersection of two sets are as follows:

	Union	Intersection
Idempotent:	$A \cup A = A$	$A \cap A = A$
Commutative:	$A \cup B = B \cup A$	$A \cap B = B \cap A$
Associative:	$A \cup (B \cup C) = (A \cup B) \cup C$	$A \cap (B \cap C) = (A \cap B) \cap C$

It should be noted that, in general,

$$(A \cup B) \cap C \neq A \cup (B \cap C).$$

Definition 1.2.6. The **symmetrical difference** of two sets A and B is $A \triangle B = \{x \mid x \in A$, or $x \in B$, but not both$\}$. The symmetrical difference of two sets is also called the **Boolean sum** of the two sets.

Definition 1.2.7. Two sets A and B are said to be **disjoint** if they do not have a member in common, that is to say, if $A \cap B = \emptyset$.

We can easily show the following theorems from the definitions of union, intersection, and complement.

Theorem 1.2.1. (Distributive Laws). Let A, B, and C be three sets. Then,

$$C \cap (A \cup B) = (C \cap A) \cup (C \cap B),$$
$$C \cup (A \cap B) = (C \cup A) \cap (C \cup B).$$

Theorem 1.2.2. (DeMorgan's Laws). Let A and B be two sets. Then,

$$\overline{(A \cup B)} = \overline{A} \cap \overline{B},$$

$$\overline{(A \cap B)} = \overline{A} \cup \overline{B}.$$

It is often helpful to use a diagram, called a Venn diagram [after John Venn (1834–1883)], to visualize the various properties of the set operations. The universal set is represented by a large rectangular area. Subsets within this universe are represented by circular areas. A summary of set operations and their Venn diagrams is given in Figure 1-1.

DeMorgan's laws can be established from the Venn diagram. If the area outside A represents $\overline{A}$ and the area outside B represents $\overline{B}$, the proof is immediate.

Let U be our universe; applying DeMorgan's laws, $A \cup B$ can be expressed as a union of disjoint sets:

$$A \cup B = \overline{(\overline{A} \cap \overline{B})} = U - (\overline{A} \cap \overline{B}) = (A \cap B) \cup (A \cap \overline{B}) \cup (\overline{A} \cap B).$$

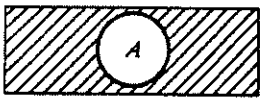
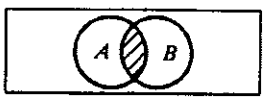
Set Operation	Symbol	Venn Diagram
Set B is contained in set A	$B \subset A$	
The absolute complement of set A	$\overline{A}$	
The relative complement of set B with respect to set A	$A - B$	
The union of sets A and B	$A \cup B$	
The intersection of sets A and B	$A \cap B$	
The symmetrical difference of sets A and B	$A \Delta B$	

Figure 1-1. Venn diagram of set operations.

Example 1.2.2.

$$
\begin{aligned}
A - (A - B) &= A - (A \cap \overline{B}) && \text{(by definition of } A - B\text{)},\\
&= A \cap \overline{(A \cap \overline{B})} && \text{(by definition of } A - B\text{)},\\
&= A \cap (\overline{A} \cup B) && \text{(by DeMorgan)},\\
&= (A \cap \overline{A}) \cup (A \cap B) && \text{(by distributive law)},\\
&= \varnothing \cup (A \cap B) && \text{(by } A \cap \overline{A} = \varnothing\text{)},\\
&= A \cap B && \text{(by } \varnothing \cup X = X\text{)}.
\end{aligned}
$$

Clearly, the elements of a set may themselves be sets. A special class of such sets is the **power set.**

Definition 1.2.8. Let A be a given set. The **power set** of A, denoted by $\mathcal{P}(A)$, is the family of sets such that $X \subseteq A$ iff $X \in \mathcal{P}(A)$. Symbolically, $\mathcal{P}(A) = \{X \mid X \subseteq A\}$.

Example 1.2.3. Let $A = \{a,b,c\}$. The power set of A is as follows:

$$\mathcal{P}(A) = \{\{\varnothing\},\{a\},\{b\},\{c\},\{a,b\},\{b,c\},\{c,a\},\{a,b,c\}\}.$$

Exercises for Section 1.2

1. List the elements in the following sets.
   (a) The set of prime numbers less than or equal to 31.
   (b) $\{x \mid x \in \mathbb{R} \text{ and } x^2 + x - 12 = 0\}$, where $\mathbb{R}$ represents the set of real numbers.
   (c) The set of letters in the word $S\ U\ B\ S\ E\ T\ S$.
2. Russell's paradox: Show that set K, such that $K = \{S \mid S \text{ is a set such that } S \notin S\}$, does not exist.
3. Prove that the empty set is unique.
4. Cantor's paradox: Show that set A, such that $A = \{S \mid S \text{ is a set}\}$, does not exist.
5. Let $U = \{1,2,3,4,5\}$, $A = \{1,5\}$, $B = \{1,2,3,4\}$, and $C = \{2,5\}$. Determine the following sets.
   (a) $A \cap \overline{B}$.
   (b) $A \cup (B \cap C)$.
   (c) $(A \cup B) \cap (A \cup C)$.
   (d) $\overline{(A \cap B)} \cup \overline{(B \cup C)}$.
   (e) $\overline{A} \cup \overline{B}$.

6. Let A, B, and C be subsets of U. Prove or disprove:

$$(A \cup B) \cap (B \cup \overline{C}) \subset A \cap \overline{B}.$$

7. Use DeMorgan's laws to prove that the complement of

$$(\overline{A} \cap B) \cap (A \cup \overline{B}) \cap (A \cup C)$$

is

$$(A \cup \overline{B}) \cup (\overline{A} \cap (B \cup \overline{C})).$$

8. A_k are sets of real numbers defined as

$$A_0 = \{a \mid a \leq 1\}$$

$$A_k = \{a \mid a < 1 + 1/k\}, k = 1, 2, \ldots.$$

Prove that

$$\bigcap_{k=1}^{\infty} A_k = A_0.$$

9. List the elements of the set $\{a/b$: a and b are prime integers with $1 < a \leq 12$ and $3 < b < 9\}$.
10. Let A be a set. Define $\mathcal{P}(A)$ as the set of all subsets of A. List $\mathcal{P}(A)$, where $A = \{1,2,3\}$. If $\mathcal{P}(A)$ has 256 elements, how many elements are there in A?
11. If set A has k elements, formulate a conjecture about the number of elements in $\mathcal{P}(A)$.
12. The **Cartesian product** of the sets S and T, $(S \times T)$, is the set of all ordered pairs (s,t) where $s \in S$ and $t \in T$, with $(s,t) = (u,v)$ for $u \in S$, $v \in T$, iff $s = u$ and $t = v$. Prove that $S \times T$ is not equal to $T \times S$ unless $S = T$ or either S or T is $\varnothing$.
13. Prove that $B - A$ is a subset of $\overline{A}$.
14. Prove that $B - \overline{A} = B \cap A$.
15. Prove that $A \subset B$ implies $A \cup (B - A) = B$.
16. If $A = \{0,1\}$ and $B = \{1,a\}$, determine the sets
    (a) $A \times \{1\} \times B$.
    (b) $(B \times A) \times (B \times A)$.

Example 4.7.3. Let S be the set $\{0,1\}$ and let $\oplus$ and $\otimes$ be the operators OR and AND defined by the table:

x	y	x OR y	x AND y
0	0	0	0
0	1	1	0
1	0	1	0
1	1	1	1

Let A be the matrix in Example 4.7.1 and B be the matrix in Example 4.7.2. The inner product A OR.AND B is the matrix

$$\begin{bmatrix} 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 0 & 1 & 1 \\ 1 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

We will show how two representative elements of this matrix are obtained. The entry in row 2, column 2 is obtained from the second row of A and the second column of B, and is (0 AND 1) OR (1 AND 0) OR (1 AND 0) OR (1 AND 1) OR (1 AND 0) = 1. The entry in row 2, column 3 is obtained from the second row of A and the third column of B, and is (0 AND 1) OR (0 AND 1) OR (1 AND 1) OR (0 AND 0) OR (0 AND 0) =

Definition 4.7.4. When the two operations $\oplus$ and $\otimes$ are the particular operations of OR and AND, respectively, then we shall refer to their inner product as the **Boolean product.**

The following theorem expresses an important computational relationship between binary relations and these operations on Boolean matrices.

Theorem 4.7.1. Let R_A and R_B be binary relations on a set $\{v_1,\ldots,v_n\}$, represented by adjacency matrices A and B respectively. Then the Boolean product A OR.AND B is the adjacency matrix of the relation $R_A \cdot R_B$, and the matrix $(\text{OR.AND})^n A$ is the adjacency matrix of the relation R_A^n. Here $(\text{OR.AND})^2 A$ means A OR.AND A.

Proof. Recall that $R_A \cdot R_B = R_C$ where $R_C = \{(v_i,v_k) \mid (v_i,v_j) \in R_A$ and $(v_j,v_k) \in R_B$ for some $j\}$. Thus if C is the adjacency matrix of R_C

$C(i,k) = 1$ iff for some j, $A(v_i,v_j) = 1$ and $A(v_j,v_k) = 1$. This is exactly the same as saying $C(i,k) = (A(i,1)$ AND $B(1,j))$ OR ... OR $(A(i,n)$ AND $B(n,j)$, which is $(A$ OR.AND $B)(i,j)$. The theorem is thus a direct consequence of the definitions. □

Corollary 4.7.1. Let A be the adjacency matrix of any binary relation R on a set $V = \{v_1,\ldots,v_n\}$. Then the adjacency matrix of the transitive closure R^+ is given by A OR $(\text{OR.AND})^2A$ OR ... OR $(\text{OR.AND})^nA$.

Proof. This follows from the preceding theorem and the fact that $R^+ = R \cup R^2 \cup \cdots \cup R^n$. □

Corollary 4.7.2. Let A be the adjacency matrix of any finite binary relation R. The adjacency matrix of the transitive reflexive closure of R, R^*, is given by $(I$ OR A OR $(\text{OR.AND})^2$ A OR ... OR $(\text{OR.AND})^n$ $A)$, where I is the identity matrix

$$I = \begin{bmatrix} 1 & 0 & \cdots & 0 \\ 0 & 1 & & 0 \\ \vdots & & \ddots & \vdots \\ 0 & 0 & \cdots & 1 \end{bmatrix}.$$

Proof. This is left to the reader as an exercise. □

Example 4.7.4. Let us apply Corollary 4.7.1 to the relation R of Example 4.6.2. In other words, $A = \{a,b,c,d\}$ and $R = \{(a,a)(a,b), (b,c),(c,d),(c,e),(d,e)\}$. Let A_R^k denote the adjacency matrix of the relation R^k. By Theorem 4.7.1, A_R^2 is the Boolean product of A_R and A_R, that is, $A_R^2 = A_R$ OR.AND A_R. Likewise, for $k > 1$, A_R^k is the Boolean product of A_R^{k-1} and A_R.

We see that

$$A_R = \begin{bmatrix} 1 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix} \qquad A_R^2 = \begin{bmatrix} 1 & 1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

$$A_R^3 = \begin{bmatrix} 1 & 1 & 1 & 1 & 1 \\ 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix} \qquad A_R^4 = A_R^5 = \begin{bmatrix} 1 & 1 & 1 & 1 & 1 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}.$$

Of course, we can compute the separate matrices by appeal to the graph and observing paths of different lengths. For instance, the second row of A_R^2 means that there is a path of length 2 from b to d and one of length 2 from b to e. The third row of A_R^2 reflects the fact that there is a path of length 2 from c to e.

Now if $\vee$ stands for the Boolean operation OR, then the adjacency matrix of the transitive closure R^+ is:

$$A_R{}^+ = A_R \vee A_R^2 \vee A_R^3 \vee A_R^4 \vee A_R^6 = \begin{bmatrix} 1 & 1 & 1 & 1 & 1 \\ 0 & 0 & 1 & 1 & 1 \\ 0 & 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

Making use of operations other than "AND" and "OR", and integer matrices, it is possible to extract other useful information from an adjacency matrix. Among such useful information is the number of distinct paths of a given length from one vertex to another in a digraph, the length of the shortest path between two vertices, and the length of a longest path between two vertices.

Theorem 4.7.2. Suppose $G = (V,E)$ is a directed graph and A is its adjacency matrix. Let $\oplus$ and $\otimes$ denote the operations

$$x \oplus y = \begin{cases} x \text{ if } x > y \\ y \text{ otherwise,} \end{cases}$$

$$x \otimes y = \begin{cases} x + y \text{ if } x > 0 \text{ and } y > 0 \\ 0 \text{ otherwise,} \end{cases}$$

Phys. Med. Biol. 44 (1999) R99–R155. Printed in the UK PII: S0031-9155(99)36024-3

TOPICAL REVIEW

Dose calculations for external photon beams in radiotherapy

Anders Ahnesjö† and Maria Mania Aspradakis‡

† Helax AB, Box 1704, Klostergatan 12, 751 47 Uppsala, Sweden
‡ Regional Medical Physics Department, Newcastle General Hospital, Newcastle upon Tyne, NE4 6BE, UK

E-mail: anders.ahnesjo@helax.se and maria-mania.aspradakis@ncl.ac.uk

Received 24 November 1998, in final form 15 July 1999

Abstract. Dose calculation methods for photon beams are reviewed in the context of radiation therapy treatment planning. Following introductory summaries on photon beam characteristics and clinical requirements on dose calculations, calculation methods are described in order of increasing explicitness of particle transport. The simplest are dose ratio factorizations limited to point dose estimates useful for checking other more general, but also more complex, approaches. Some methods incorporate detailed modelling of scatter dose through differentiation of measured data combined with various integration techniques. State-of-the-art methods based on point or pencil kernels, which are derived through Monte Carlo simulations, to characterize secondary particle transport are presented in some detail. Explicit particle transport methods, such as Monte Carlo, are briefly summarized. The extensive literature on beam characterization and handling of treatment head scatter is reviewed in the context of providing phase space data for kernel based and/or direct Monte Carlo dose calculations. Finally, a brief overview of inverse methods for optimization and dose reconstruction is provided.

Contents

0031-9155/99/110099+57$30.00

1. Introduction

In the early days of radiotherapy, 'dose' was used in a pharmacological sense quantifying the amount of radiation given rather than its physical impact on the irradiated matter. Today, absorbed dose is strictly defined as mean energy imparted (by ionizing radiation) per mass (ICRU 1998), i.e. dose is decoupled from the radiation used to deliver it. Therefore, absorbed dose is the fundamental physical quantity of interest for relating radiation treatment to its outcome. The broad spectrum of events that impart energy to matter implies that direct measurement of dose from its definition is not a trivial task. Calorimeters and ionization chambers can be used to measure dose in absolute terms but are not suitable for *in vivo* dosimetry. Thermoluminescence detectors and diodes, placed on the patient surface or within cavities, are used to check the delivered dose in patients, but they are not suitable for obtaining a map of the dose. Hence, quantification of dose distributions in patients must be based on calculation models, both for treatment planning and in following up the delivered treatment.

When high-energy photon beams were introduced clinically in the 1950s, the only choice available for medical physicists was to develop empirical methods. These methods often restricted treatments to beam set-ups that could be calculated with some confidence. A historical review of treatment techniques with an extensive bibliography of early developments in dose calculations is provided by Fraass (1995). Before access to CT scanning, mapping of patient anatomy had to rely on simple contouring, i.e. dose calculation accuracy was limited to a great extent by the lack of relevant information about the patient. During the early 1970s 'mini-computers' provided enough computational power to enable the development of both CT scanners and computer based treatment planning systems at prices affordable to healthcare providers. These advents boosted the refinement of empirical dose calculation methods to incorporate voxel by voxel considerations of heterogeneous tissues (see reviews of that time, e.g. Purdy and Prasad (1983) or Cunningham (1983)). More recently, imaging modalities such as MRI (nuclear magnetic resonance imaging), SPECT and PET (single photon and positron emission tomography) have greatly increased the possibility of localizing and delineating tumours and nearby organs (Austin-Seymour *et al* 1995, Schad *et al* 1992). In response to greater precision in defining the target volume, a present trend is to explore all available

degrees of freedom in beam delivery in order to increase the target dose and spare normal tissues (Bortfeld *et al* 1997, Brahme 1987, 1995, Chin *et al* 1983, Webb 1997). Increased dose requires increased accuracy, as reviewed in section 3.2. Empirical methods are limited in accuracy and often fail to model generalized beam set-ups. Although the transport equation and the interaction cross sections are well known, no analytical dose calculation algorithms for photon beams have been developed that are general enough to handle radiotherapy geometries with high enough accuracy. Analytical methods might work if a very short chain of events precedes the complete absorption of the particle's energy (or when the photons escape the system of concern). Monte Carlo methods, implemented to mimic the basic processes in a straightforward way, have served many purposes in medical physics (see reviews by Andreo (1991), Rogers and Bielajew (1990) and Mackie (1990)). However, they have not yet become suitable for routine treatment planning of photon beams due to their huge requirement for CPU time. Therefore, a new family of semianalytical dose calculation algorithms based on energy deposition kernels has been developed, as reviewed in detail in section 7.

Traditionally, calculation of dose output and the related irradiation time or the accelerator monitor units has been treated as a separate task, often not integrated into the treatment planning system itself. More recently, attention has been drawn to the importance of characterizing the beam and to fully modelling output and lateral beam variations. The result is a more complete understanding and modelling of dose deposition, thus enabling planning to be carried out for more complicated treatments. Increased requirements on standards for safety and quality assurance during treatment have, on the other hand, emphasized the important role of simple dose calculation methods for independent checks of the output from treatment planning systems.

1.1. The scope of this review

The aim of this work is to review the background, requirements, formalisms and algorithms for photon beam dose modelling in external radiotherapy. Calculation methods for brachytherapy and radiation protection fall outside the scope of this review. The emphasis will be on methods suitable for the implementation and/or check of dose calculations in 3D treatment planning systems. Beginning with introductory sections on the energy deposition processes and clinical requirements for dose calculations, we continue with formalisms for monitor unit normalization followed by several sections on particular dose calculation methods. The methods will be described in order of increasing explicitness of particle transport considerations. The range of modelling starts with empirically oriented techniques (sections 5 and 6) and continues over kernel based methods (section 7) to explicit particle transport methods such as Monte Carlo (section 8). There is, however, no strict division between models based on first-principle particle transport and empirical models. A particular implementation may consist of elements from both groups of models and also apply different monitor unit calibration/normalization formalisms. All models require the incident beam to be characterized to provide basic data as discussed in section 9. Finally, some inverse techniques proposed for beam optimization are reviewed in section 10.

2. Energy deposition in photon beams

2.1. The physical processes

The photons from a treatment machine yield a cascade of interactions, not only in the patient but also in the treatment machine itself before the energy is absorbed as dose (see figure 1).

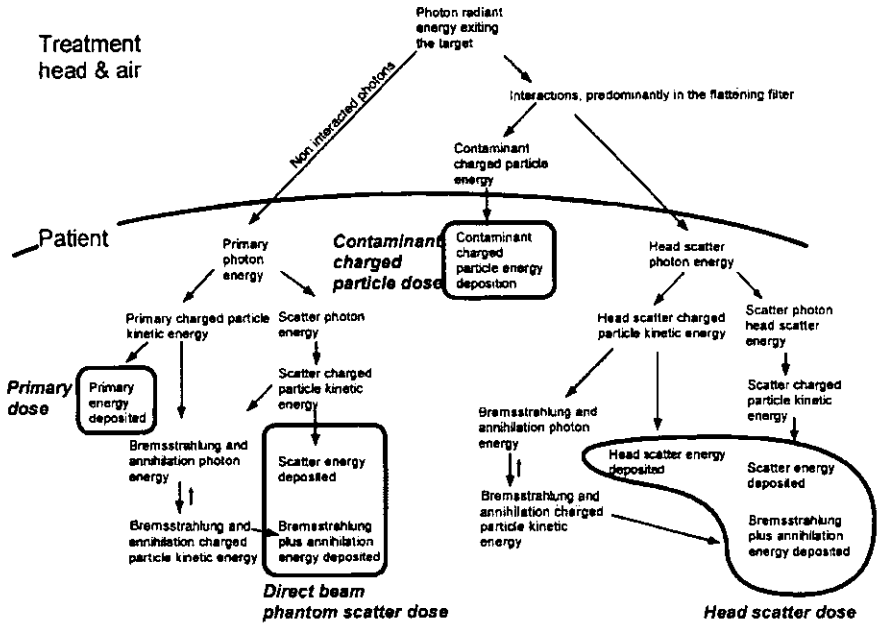


Figure 1. Interaction history of the four dose categories commonly referred to in dose calculations for treatment planning—primary dose, phantom scatter dose, contaminant charged particle dose and head scatter dose.

Following irradiation of the treatment head elements, the beam is scattered adding a secondary photon component to the primary beam. In addition, charged particles released in the treatment machine head and the air column between the head and the irradiated medium contaminate the beam and contribute to the dose in the build-up region. The amount of charged particle contamination is very sensitive to the presence of scattering material. Therefore, dosimetry protocols state that beams should be calibrated at a depth beyond the range of charged particle contamination (IAEA 1987).

Photons are indirectly ionizing particles and do not deposit significant energy themselves. Through interactions with atoms in the patient, the incident photons transfer their energy to electrons and positrons that ionize and excite atoms along particle tracks until their energy is lost. Using the interaction history one can make unambiguous definitions of the various dose categories relevant to beam characterization and dose modelling, as outlined in figure 1. Starting at the source (electron beam target), most photons entering the patient have not been subject to any interactions before entering the patient and will serve as originators of the primary and phantom scatter dose distributions. Particles interacting in the treatment head yield two dose categories: charged particle contamination and head scatter dose. The different order of scatter regarding the head scatter dose is not normally separated. Head scatter dose accounts, depending on beam energy, for approximately 5–15% of the total dose (Ahnesjö 1994).

Since the time of conventional x-rays and ^{60}Co units the importance of charged particle transport has often been overlooked in dose calculations for treatment planning. It has been considered sufficient to assume that the photon energy transferred to such particles was deposited 'on the spot' (collision kerma approximation). This has caused confusion, for instance, when 'primary dose' has been defined experimentally as 'zero-area tissue–phantom

ratio' (Mohan and Chui 1985). However, it is convenient to keep the concept of primary dose since its dependence on the primary photon fluence is far more local than the dose mediated by scattered photons (see Bjärngard and Cunningham 1986, Nizin 1993 and Woo *et al* 1990). In ICRU (1987), primary radiation is taken to be the radiation incident on the surface of the phantom and includes photons coming directly from the target as well as radiation scattered from the beam shaping and collimating system. In this review, unless otherwise stated, dose due to radiation that has been scattered within the head of the treatment machine (resulting in head scatter dose) will be considered separately. Reasons for the separation are that head scattering processes are independent of scattering in the patient and results in radiation that differs in energy and direction from the primary beam. In section 9, the literature on head scattered radiation is reviewed in more detail.

2.2. The theorems according to Fano and O'Connor

The dosimetric data used in treatment planning are mainly derived for water. The existence of two important theorems by Fano and O'Connor enables density-scaling of data for water to 'water-like media' with arbitrary densities.

Fano's theorem states that when an object of varying density but constant atomic composition is present in a radiation field with constant fluence of primary particles (photons), then the fluence of secondary particles (electrons) is also constant and independent of the density variations (Fano 1954). This constant fluence of secondary electrons equals the fluence in charged particle equilibrium (CPE) for a given fluence of photons. Consequently the absorbed dose across any area of density variations would be constant. The main assumption in Fano's theorem is that the interaction cross sections per unit mass are independent of the density of a medium of identical atomic composition. Strictly, in order to apply Fano's theorem to external photon beams, one must assume that primary photon attenuation, the stopping power density effect and the release of secondary photons can be neglected. Ignoring photon attenuation essentially means that the mean free paths of primary photons must be much larger than the maximum ranges of the released secondary electrons. This first condition can be fulfilled in clinical beams, with photon energies less than 1–3 MeV and applies to points in an externally irradiated medium which are sufficiently far from boundaries (Harder 1974). Density effects (within the density range of human tissues) are generally small for clinical beams and the production of secondary photons is not problematic as long as their mean free paths are larger than the ranges of secondary electrons. For the above reasons Fano's theorem is an important test of dose calculation algorithms (Nilsson and Knöös 1992). The effect of lateral charged particle disequilibrium is illustrated in figure 2 by the depth dose curves along the central axis in fields of different sizes.

While Fano's theorem applies to situations of charged particle equilibrium, the density scaling theorem by O'Connor relates the dose in two media of different density but equal atomic composition, both irradiated by the same external beam, to each other. According to this theorem, the ratio of the secondary scattered photon fluence to that of primary photon fluence is constant in the two media provided all geometric distances, including field sizes, are scaled inversely to the density (O'Connor 1957). This means that the dose at corresponding points in two media is the same if all dimensions in the irradiation geometries are scaled inversely with density (see figure 3). Both Fano's and O'Connor's theorems rely on a common assumption that the interaction probability (per electron) is independent of density variations between media. The common foundations and relations between these two theorems was analysed by Bjärngard (1987).

 3256

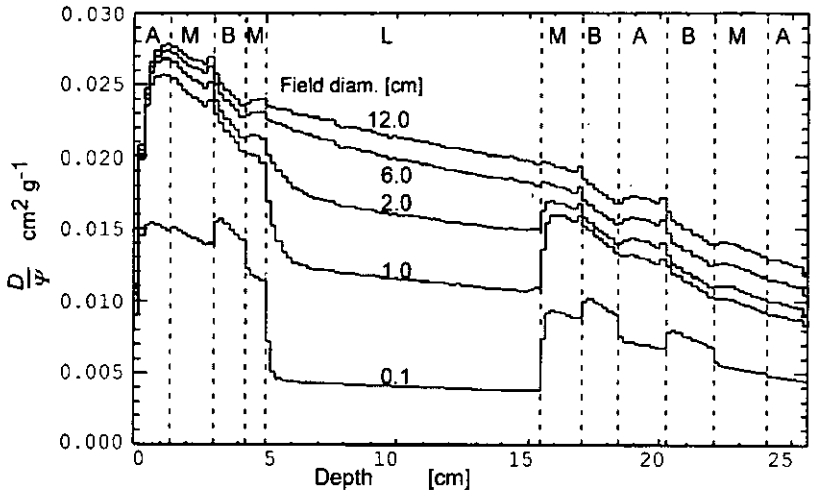


Figure 2. Central axis depth dose curves calculated with the Monte Carlo method for parallel (infinite SSD) 4 MV circular beams of varying diameters (ranging from 0.1 to 12.0 cm) onto a stack of tissue media composed of adipose (A), muscle (M), bone (B) and lung (L) with densities 0.92, 1.04, 1.85 and 0.25 g cm^{-3} respectively. For small fields there is a great difference in dose to different media because of the greatly varying degree of lateral equilibrium of the released charged particles. At larger field sizes the dose is rather constant and declines with depth according to the attenuation of the primary beam. (Adapted from Ahnesjö (1987,1989).)

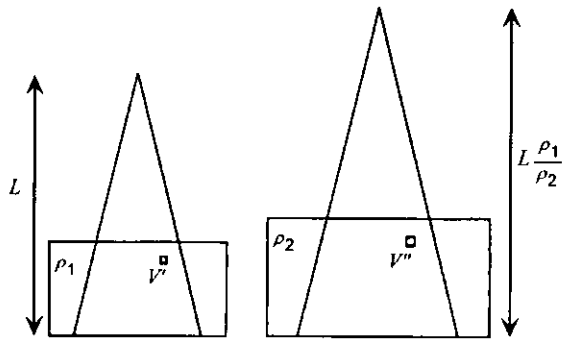


Figure 3. The dose at V' and V'' are equal according to O'Connor's theorem provided that all linear dimensions (including the source to surface distance) are scaled by the phantom density and the number of photons per unit solid angle is equal.

2.3. The reciprocity theorem

For radiation transfer the reciprocity theorem states that *reversing the positions of a point detector and a point isotropic source within an infinite homogeneous medium does not change the amount of radiation detected* (Attix 1986). This theorem dates back to King (1912) who formulated it as a reciprocal relationship between the primary radiation from a point source and a finite volume. Mayneord (1945) extended the theorem to the case where the source and the detector are both extended: *the integral dose throughout any volume whatever*

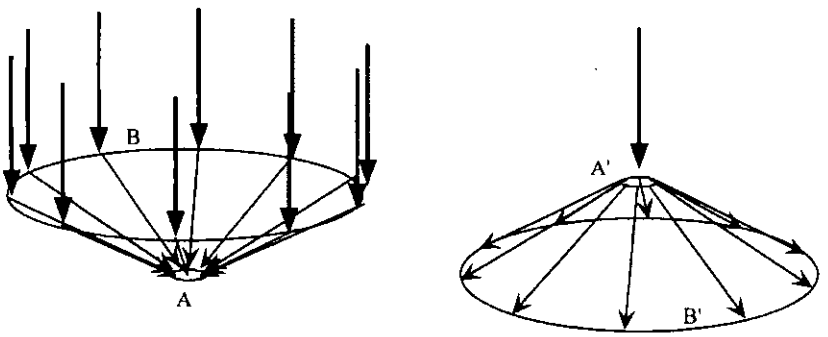


Figure 4. The estimate of dose originating from a large scattering element B to a small element A, as shown to the left, can be made with better scoring statistics by use of a reciprocal geometry as to the right in which A′ is used as the scattering element and B′ as the tally region. The set-up requires a homogeneous medium and rotational symmetric scattering around the primary particle direction for the reciprocity to apply. (Adapted from Hoban *et al* (1994), with permission.)

due to a finite source uniformly filled with radioactive material is equal to the integral dose throughout the original source if the receiver be filled with radiating material of the same uniform density. The implicit approximation in kernel superposition models (see section 7) is that kernels are treated as if the medium were infinite and homogeneous so the theorem applies directly to superposition integrals without requiring further approximations. Although kernels are generally derived assuming an infinite medium, this is not an absolute requirement for superposition calculations. Strictly kernels could be derived at boundary positions and used for superposition calculations (Woo 1994). Kernel reciprocity means that point dose kernels as well as describing the energy deposited around a photon interaction site, also describe the distribution of scattering sites for the particles that mediate energy to a dose deposition point. Due to this reciprocity, comparison between calculated energy spread kernels and measured *iso-line* dose contribution curves was possible (O'Connor and Malone 1989). The application of the theorem in the case of polyenergetic beams where polyenergetic kernels are employed (see section 7.2.1.1) is not exact because of differences in the differential energy fluence spectrum used to define polyenergetic kernels (Papanikolaou *et al* 1993). The reciprocity between photon interaction and dose deposition sites is appropriate for designing geometries for use in experiments or Monte Carlo simulations where signal to noise ratios or statistical uncertainties are of concern (see figure 4).

The reciprocity is also sometimes utilized in radiation transport codes in *adjoint* mode where particles are treated as going backwards from the tally region towards the sources (Difilippo 1998, Wagner *et al* 1994). The technique is best suited to problems where one wants to estimate the response of a small detector exposed to large distributed sources.

2.4. Common experimental quantities

Apart from a summary given as an appendix in ICRU (1976), most of the quantities traditionally used in photon beam characterization lack formal definitions approved by an international body of standardization. Several common quantities have been reviewed in a recent formalism proposal (Dutreix *et al* 1997) with the aim to serve in 'manual' calculations of monitor unit settings. In the present section we give a brief overview of the most commonly used quantities defined for points along the central axis of a photon beam. These are classified into those

 3258

quantities that express the depth penetration characteristics of beams, those that attempt to separate scatter dose from primary dose and those that describe the output of the clinical accelerator. The most important of the reviewed quantities are summarized in table 1.

To describe the penetration characteristics, three quantities have been widely used: the percentage depth dose (PDD), the tissue phantom ratio (TPR) and the tissue maximum ratio (TMR). PDD at a point in water is defined as the ratio of absorbed dose at that point to the absorbed dose at the depth of maximum build-up along the central axis. PDD data are impractical for direct reconstruction of dose distributions since they depend on the source to surface distance (SSD). Instead TPR, being independent of SSD, has gained popularity. TPR is defined as the ratio of the total absorbed dose on the central axis at depth to that at a point also on the central axis and at the same distance from the source but with the surface of the phantom moved so that the point is at a specified reference depth (Karzmark *et al* 1965). The TPR was defined to comply with recommendations that x-ray beams should be calibrated at a reference depth in a phantom (ICRU 1963). Another quantity, the tissue maximum ratio (TMR), has been used in some dosimetry systems (Holt *et al* 1970). TMR is renormalized TPR such that the specified reference depth is the depth of maximum dose. The uncertainties due to electron contamination at the depth of dose maximum is a complication, and the use of a reference depth further away from the build-up region, as in the TPR definition, is strictly a better choice for dosimetry systems (Dutreix *et al* 1997).

Early dosimetric systems have tried to separate scatter from primary dose using scatter factors to express the ratio of total to primary dose at a point (ICRU 1973). Tissue air ratio (TAR) was defined as the ratio of the absorbed dose at a given point in a phantom to the absorbed dose at the same point in air, but at the centre of a small volume of phantom material, of mass just large enough to provide electronic equilibrium, at the point of measurement (ICRU 1973, 1976). This definition of TAR (originally known as tumour air ratio by Johns *et al* (1953)), has been a subject of controversy for high-energy beams due to experimental problems in ensuring ideal charged particle equilibrium in air. In a later definition, TAR is relative to the primary

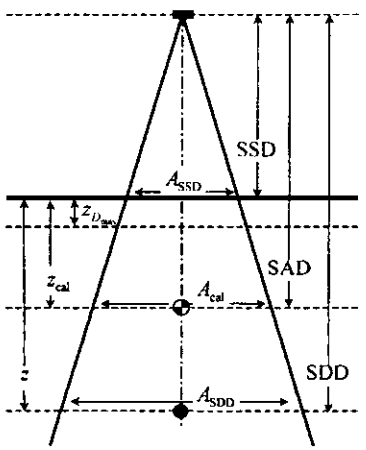


Figure 5. Geometry for specifying some of the quantities given in table 1. The calibration depth z_{cal} is often at isocentre. The acronyms SSD, SAD and SDD read source to surface distance, source to axis distance, and source to detector distance. The generic aperture variable A may, depending on context, represent the entire beam set-up geometry rather than just field size.

 3259

Table 1. Some of the most common quantities used in photon beam characterization. The most common abbreviation and symbols in each context have been used (cf figure 5), although some arguments are adapted to follow the nomenclature of this review. For some quantities equivalent concepts have been defined based on the energy fluence formalism adapted from Ahnesjö (1994), Kim *et al* (1998), and section 4.1 (see that section for definitions of variables).

Quantity	Definition	Comment
Percentage depth dose	$\mathrm{PDD}(A_{\mathrm{SSD}}; z) = \frac{D(A_{\mathrm{SSD}}; z)}{D(A_{\mathrm{SSD}}; z_{\max})} \times 100$	Detector scanned in a fixed SSD beam-phantom system. Aperture size defined at SSD
Tissue phantom ratio	$\mathrm{TPR}(A_{\mathrm{SDD}}; z, z_{\mathrm{cal}}) \equiv \frac{D(A_{\mathrm{SDD}}, \mathrm{SDD}; z)}{D(A_{\mathrm{SDD}}, \mathrm{SDD}; z_{\mathrm{cal}})}$	Phantom moved in a fixed SDD beam-detector system. Aperture size defined at SDD
Tissue air ratio	$\mathrm{TAR}(A_{\mathrm{SDD}}; z) = \frac{D(A_{\mathrm{SDD}}, \mathrm{SDD}; z)}{D_{\text{build-up cap}}(A_{\mathrm{SDD}}, \mathrm{SDD})}$	Phantom moved in a fixed SDD beam-detector system. Denominator obtained 'in air' using a build-up cap on the detector
Scatter air ratio	$\mathrm{SAR}(A_{\mathrm{SDD}}; z) = \mathrm{TAR}(A_{\mathrm{SDD}}; z) - \mathrm{TAR}(A \to 0; z)$	Several proccdures proposed to determine $\mathrm{TAR}(A \to 0; z)$
Scatter primary ratio	$\mathrm{SPR}(A_{\mathrm{SDD}}; z) = \frac{\mathrm{SAR}(A_{\mathrm{SDD}}; z)}{\mathrm{TAR}(A \to 0; z)}$	
Output factor	$S_{\mathrm{cp}}(A) = \frac{D(A; z_{\mathrm{cal}})}{D(A_{\mathrm{cal}}; z_{\mathrm{cal}})}$	Definitions using energy fluence formalism: $S_{\mathrm{cp}}(A) = \frac{D(\Psi_0+\Psi_{\mathrm{hsc}}(A); A; z_{\mathrm{cal}})}{D(\Psi_0+\Psi_{\mathrm{hsc}}(A_{\mathrm{cal}}); A_{\mathrm{cal}}; z_{\mathrm{cal}})} \frac{(1+b(A_{\mathrm{cal}}))}{(1+b(A))}$
Output factor in water		
Output ratio in water		
Collimator scatter factor	$S_{\mathrm{c}}(A) = \frac{D_{\text{mini-phantom}}(A; z_{\mathrm{cal}})}{D_{\text{mini-phantom}}(A_{\mathrm{cal}}; z_{\mathrm{cal}})}$	$S_{\mathrm{c}}(A) = \frac{\Psi_0+\Psi_{\mathrm{hsc}}(A)}{\Psi_0+\Psi_{\mathrm{hsc}}(A_{\mathrm{cal}})} \frac{(1+b(A_{\mathrm{cal}}))}{(1+b(A))}$
Head scatter factor		
Output factor in air		
Mini-phantom output ratio		
Phantom scatter factor	$S_{\mathrm{p}}(A) = \frac{S_{\mathrm{cp}}(A)}{S_{\mathrm{c}}(A)}$	$S_{\mathrm{p}}(A) = \frac{D(\Psi_0; A; z_{\mathrm{cal}})}{D(\Psi_0; A_{\mathrm{cal}}; z_{\mathrm{cal}})} \approx \frac{D(\Psi_0+\Psi_{\mathrm{hsc}}(A_{\max}); A; z_{\mathrm{cal}})}{D(\Psi_0+\Psi_{\mathrm{hsc}}(A_{\max}); A_{\mathrm{cal}}; z_{\mathrm{cal}})}$
Volume scatter ratio		

dose at the depth of maximum dose build-up, still with constant SDD (BJR 1983 (also in BJR 1996)).

Scatter primary ratio, SPR, is a general name for quantities defined as ratios of scatter versus primary dose at a point. A strict definition for SPR is given by Bjärngard and Petti (1988) where the denominator represents dose originating from primary photons at a depth larger than that of maximum build-up (cf section 4.2). A related quantity is scatter air ratio (SAR) describing the absorbed dose originating from scattered radiation, and is practically derived by subtracting the extrapolated value for zero-area field TAR from finite field TAR (Gupta and Cunningham 1966). The use of extrapolation, however, is ambiguous in strictly representing primary dose (Kijewski *et al* 1986). SAR can be differentiated with respect to field radius to yield differential scatter air ratios (dSAR) (Cunningham 1972, Larson and Prasad 1978, Sontag and Ray 1995). Analogous to SAR, scatter maximum ratios (SMR) (originally called scatter phantom ratios by Khan *et al* (1972)) are calculated by subtracting the zero-area TMR for that depth from the TMR at the same depth and beam radius. In both SAR and SMR the denominator is the primary dose at the depth of maximum build-up. Modelling dose from scattered photons using dSAR or differentiated SPR can be done using various integration techniques (see section 5.2.2). Other commonly used scatter factors are: the backscatter factor (BSF), defined as the SPR at the surface of the phantom on the central axis, and used for low- and medium-energy x-rays, and the peak scatter factor (PSF), used for higher-energy beams and defined at the depth of maximum build-up. Normalized peak scatter factors (NPSF), are PSF renormalized to a reference field size to avoid uncertainties due to scattered photons (Day 1983).

The output from a treatment machine is defined in ICRU 24 (ICRU 1976) as the rate of exposure or dose for a given field related to the same quantity in a reference geometry which usually is the calibration depth and field size. The separation of total output ('output in water') into treatment head scatter ('output in air') and phantom scatter was first done by Holt *et al* (1970). They measured the total output and head scatter factors separately, as they are readily defined, unlike phantom scatter factors which can only be measured with 'field-cut' phantoms with maximum-opened jaws on the machine. Phantom scatter factors, labelled as S_p, are often estimated as the ratio of the total scatter factor S_{cp} to the collimator (head) scatter factor S_c, thereby assuming equal broad beam dose per energy fluence conversion of primary and head scattered photons. (Other common symbols for these factors are OF_{phant}, OF_w and OF_{air} respectively.) This is not strictly true since the diffuse beam of head scattered photons is larger than the well-collimated primary beam, but considering that head scatter is only a fraction of the total the approximation can be used. The various kinds of output factors can be given strict definitions by applying energy fluence formalisms (Ahnesjö 1994, Kim *et al* 1998) (see table 1).

3. Dose calculations for treatment planning

Dose calculation models should serve, within the environment of a treatment planning system, to provide quick and accurate results for all relevant types of treatment field arrangements. The demands on dose calculations are therefore to a large degree context dependent. Important aspects in design of treatment planning systems are not only the accuracy of the results but also the logistics of the data objects, user interface layouts, etc.

3.1. Generality, flexibility and logistics

The general requirements for a treatment planning system were identified early (ICCR 1970) although the technology to realize the goals has only just started to approach the real needs.

Anatomy mapping with CT enabled the modern approach to treatment planning where much of the beam geometry is evaluated by a combination of dose quantification and 'virtual simulation' of a CT study rather than physical simulation with the patient at a simulator (Goitein and Abrams 1983). The 'beam's-eye view concept' (Goitein *et al* 1983) is a classical invention mimicking radiographs to align the beam with identified organs. This concept has now merged with dose calculations as to verify delivered dose by patient dose reconstruction from portal dose distributions (see section 10.2). The use of intensity modulation to optimize beam profiles (see Brahme (1995) for a review) requires that the beam can be modulated to achieve the desired treatment. Use of scanning beams and dynamic multileaf modulation will not affect spectral-dependent beam properties but if attenuating modulators ('compensator filters') are applied, the change in spectral-dependent properties should be accounted for. Hence, the dose model framework must be general enough to include the influence on dose and beam properties from general beam set-ups and modulations, and also be able to follow the particle transport to the portal image plane with adequate details.

Jung *et al* (1997) have studied the clinical workflow and developed design principles to meet logistic demands. As an example it was stated that it should be possible to perform the different tasks in a natural, and as far as possible, arbitrary order. It implies that it could be desirable to know the dose to a single point before a full matrix calculation is done. Most traditional algorithms, that typically take into account only the ray path from the source to the calculation point, can generate dose values at arbitrary points. A modern approach, such as that from convolution/superposition or Monte Carlo, provides the result in a 3D Cartesian grid but calculation efficiency is lost completely if dose has to be delivered point by point. This is of some importance as the number of points where the dose is calculated can be reduced significantly if an optimal set of points is used (Niemierko and Goitein 1990). However, Cartesian grids are sufficient if the grid resolution is high enough (see the paper by van't Veld and Bruinvis (1995) which includes a bibliography). An obvious choice is to have a highly integrated 'single point model' in interactive operation where some accuracy is traded for point dose speed, and a second, more accurate, 'bulk matrix model' which could be allowed to run in batch mode. It is, however, desirable to use a single model in order to simplify clinical verification and quality assurance (although a multimodel system would provide some inherent quality assurance comparing results from the different models). For interactive use, point calculations should be virtually instantaneous and spend a maximum of some tens of seconds in calculating a dose distribution in a patient cross section. The uppermost time limit for an accurate calculation to a bulk volume when no user interaction is required would be 1 h, allowing reasonable patient throughput. Use of optimization based on biological or physical object functions will, however, increase the demand for computational speed since optimization schemes usually use iterative search methods (Gustafsson 1996). For gradient based search methods one also need to know the gradient of the response R with respect to the optimization variable ν:

$$\frac{\partial R}{\partial \nu} = \frac{\partial R}{\partial D}\frac{\partial D}{\partial \nu} \tag{1}$$

i.e. the dose model should provide also the gradient $\partial D/\partial \nu$ besides the dose D. Dose must be recomputed occasionally during the optimization search since the dose response $\partial R/\partial D$ is usually nonlinear and/or constrained. For intensity modulation, ν represents the energy fluence Ψ through a pixel (bixel) of the beam cross section. Commonly, the number of iterations needed in optimization is in the order of 50 to 100 leaving substantially less than a minute for dose calculations alone.

Well defined, standardized data objects will facilitate object oriented approaches using databases to serve treatment planning and dose calculations (NEMA 1998). The current

 1262

DICOM standard does not (yet) cover the data set needed for treatment unit characterization. This complicates the modularization of dose calculations into exchangeable modules using different models. Quality assurance on dose and monitor unit settings in treatment planning requires checks of the entire chain from beam characterization measurements to final output calculations. Although, in principle, modern algorithms using kernel superposition or direct Monte Carlo simulations should be superior to more simple, traditional approaches, there is a need for these as independent calculators. Thus, since simpler methods could serve for checking information fed into the treatment planning system as well as results generated by it, space will be devoted in this review to both types of calculations.

3.2. Accuracy

The basis for radiotherapy is that cancer cells are more sensitive to ionizing radiation than normal cells. Important parameters to describe the response are D_{50}, the 50% response dose, and the normalized dose gradient γ (Brahme 1984). For tumour control the D_{50} value increases with tumour size and for normal tissue injury D_{50} decreases with larger irradiated volumes (Källman *et al* 1992b), i.e. the therapeutic window shrinks with increasing tumour size. Values for γ vary from 1.5 to 7, i.e. the uncertainty in delivered dose amplifies between 1.5 to 7 times when viewed as dose related part of the uncertainty in biological response. In an attempt to quantify the actual accuracy needed, Boyer and Schultheiss (1988) studied the influence of dose uncertainty on complication-free tumour control (called 'utility function' in their paper) and concluded that a 1% higher accuracy results in 2% increase of cure. It is not surprising that extensive research has been targeted to develop dose response models suitable for application for optimization of dose distributions in treatment planning (see Brahme (1995) for a review). Major problems are not only the determination of the actual parameters to be used in the models (Ågren-Cronqvist 1995) but also the major foundations of the models which are at present subject to some controversy (Dasu and Denekamp 1999). To improve the state of the art, high accuracy and quality must also be enforced in dose reporting (Overgaard and Bartelink 1995, Dische *et al* 1993). Several general recommendations of dose delivery accuracy have been issued: 5% (ICRU 1976), 3.5% (Mijnheer *et al* 1987), 3% (Brahme 1988). The dosimetric error in absolute dose calibration has been determined by Andreo (1990). Excluding beam monitoring instabilities, the absolute dosimetry uncertainty is stated to be 2.0% for MV photon beams and 1.6% for ^{60}Co. Considering the complexity of the dose delivery process, it is of course difficult to achieve 3% accuracy in practice and it is common to refer to the ICRU 24 (ICRU 1976) value of 5% as the level for corrective action. A conservative approach for setting the limits for dose calculation errors alone is to identify the other errors in the dose delivery chain and vary the dose calculation error to identify the limit where the overall value is seriously affected by the dose calculation error (Ahnesjö 1991). Combining the dosimetry estimates from Andreo (1990) and delivery estimates from Brahme (1988) as a representation of the present technique indicates that dose calculations do not need to be better than 2% (see table 2) with a correction action level at 4%. It is unlikely that revolutionary accuracy improvements in dose delivery will occur in future, although some evolution should be anticipated. Developments in basic dosimetry, detector technology and accelerator stability may cut the errors in dose calibration, beam monitoring and flattening to half their present values. Patient data and beam–patient set-ups are difficult to improve but a reduction to two-thirds of their present values should be possible. Summarizing these expectations, a dose calculation accuracy of 1% will be sufficient as the ultimate future goal.

More specific requirements on commissioning and quality assurance of treatment planning systems have been worked out by Dahlin *et al* (1983), Van Dyk *et al* (1993) and Fraass

Table 2. Determination of accuracy goal in dose calculations. With present delivery and calibration technique 2–3% should be the aim while 1% might be the ultimate accuracy goal.

	Present technique $100 \times \Delta D(1\sigma)/D$	Future development $100 \times \Delta D(1\sigma)/D$
Absorbed dose determination at the calibration point	2.0	1.0
Additional uncertainty for other points	1.1	0.5
Monitor stability	1.0	0.5
Beam flatness	1.5	0.8
Patient data uncertainties	1.5	1.0
Beam and patient set-up	2.5	1.6
Overall excluding dose calculation	**4.1**	**2.4**
Dose calculation	1.0 2.0 3.0 4.0 5.0	0.5 1.0 2.0 3.0 4.0
Resulting overall uncertainty	**4.2 4.6 5.1 5.7 6.5**	**2.4 2.6 3.1 3.8 4.7**

Table 3. Dose and positional accuracy criteria for photon beam dose calculations adapted from Van Dyk *et al* (1993). Percentage figures are specified relative to the calibration value with a conversion to local values (within brackets) for the low-dose region. Positional accuracy values (fourth column) are defined as the distances between measured dose points and the nearest points in a calculated distribution which contain the same dose values (van't Veld 1997, Harms *et al* 1998).

	Central axis (except build-up)	High dose region Low dose gradient	Large dose gradient	Low dose region Low dose gradient
Homogeneous water slab—simple fields	2%	3%	4 mm	3% (~50%)
Stack of tissue slabs—simple fields	3%	3%	4 mm	3% (~50%)
Anthropomorphic phantoms—complex beams		4%	4 mm	3% (~50%)

et al (1998). Based on a general 4% requirement for test of existing systems (and a 2% recommendation as a 'developer's goal') Van Dyk *et al* identified a number of situations which they *a priori* assumed to present variable degrees of difficulty. They also differentiated the accuracy criterion for different dose regions, identifying that 3% local accuracy is almost meaningless in high-gradient regions and low-dose regions (see table 3).

4. Dose per monitor units normalization

Here we shall review three major formalisms for dose per monitor units normalization, one model driven dose-to-energy-fluence formalism, one based on dose-to-kerma relations and one empirically oriented dose-to-dose ratio formalism. This classification is similar to the approach used by Mackie *et al* (1996) in their review, although in this review the formalism aspect will be more thoroughly expressed. By 'formalism' we mean the quantities and their relations needed for a monitor unit calculation, while by 'model' we mean the calculational model used to evaluate the quantities used by the formalism. Hence, a formalism can be viewed as a framework, or 'top level' model, within which different computation models can be implemented.

4.1. Dose-to-energy fluence formalisms

The description here follows the work by Ahnesjö and co-workers (Ahnesjö *et al* 1992a, 1995, Ahnesjö 1994, 1995, Weber *et al* 1996, Åsell 1999). A similar but independent development

has also been presented by Mackie *et al* (1995). The common basic idea utilized is that dose is linear to the amount of radiation the patient is exposed to. The linearity makes it natural to express the dose calculated from 'first principle' models as dose per energy fluence, i.e. a dose calculation 'engine' is supposed to deliver

$$d(x, y, z) = \frac{D(x, y, z|\Psi(A; x, y, z_0))}{\Psi_0} \tag{2}$$

where Ψ_0 is a reference energy fluence level, A is a general aperture variable representing all beam collimating and modulating elements and $D(x, y, z|\Psi(A; x, y, z_0))$ is the absorbed dose at point (x, y, z), given that the lateral energy fluence distribution $\Psi(A; x, y, z_0)$ is defined free in air at a reference distance z_0 from the source. Beam attenuation and divergence are assumed to be intrinsic parts of the dose calculation model and not part of the formalism itself. A natural definition of the reference energy fluence Ψ_0 is the primary (cf figure 1) unscattered energy fluence free in air at the calibration point, normally the isocentre (Ahnesjö 1994). Lateral variations of the primary energy fluence (at z_0) are then related to Ψ_0 by the relative distribution $f(A; x, y, z_0)$ such that $\Psi_{\text{prim}}(A; x, y, z_0) = \Psi_0 f(A; x, y, z_0)$. Adding scattered photons, Ψ_{hsc}, from irradiated parts of the treatment head then yields the total photon energy fluence of the beam as

$$\Psi(A; x, y, z_0) = \Psi_0 \left(f(A; x, y, z_0) + \frac{\Psi_{\text{hsc}}}{\Psi_0}(A; x, y, z_0) \right). \tag{3}$$

Following Ahnesjö *et al* (1992a), the monitor units registered for a given beam are separated in two parts, M_0 and M_b. M_0 is the signal proportional to the forward fluence through the monitor chamber and $M_b = M_b(A)$ is proportional to the fluence of particles backscattered into the monitor from the upper part of the adjustable collimators. The backscatter signal is usually small, i.e. $b(A) = M_b(A)/M_0 \ll 1$. The total energy fluence delivered free in air per monitor unit now follows as

$$\frac{\Psi(A; x, y, z_0)}{M} = \frac{\Psi_0}{M_0}(1 + b(A))^{-1} \left(f(A; x, y, z_0) + \frac{\Psi_{\text{hsc}}}{\Psi_0}(A; x, y, z_0) \right). \tag{4}$$

The link between monitor units and energy fluence is provided by a dose normalization for the calibration geometry of the treatment unit. Combining equations (2) to (4) and simplifying the notation of absorbed dose D yields

$$\frac{D(A; x, y, z)}{M} = \frac{\Psi_0}{M_0}(1 + b(A))^{-1} d(A; x, y, z). \tag{5}$$

By requiring the measured dose, for a calibration field A_{cal} and position $(x_{\text{cal}}, y_{\text{cal}}, z_{\text{cal}})$, to equal the calculated dose for the same conditions, Ψ_0/M_0 follows as the ratio between a measured dose (per monitor unit, corrected for monitor backscatter) and a calculated dose (per energy fluence) for the calibration conditions:

$$\frac{\Psi_0}{M_0} = \frac{[D(A_{\text{cal}}; x_{\text{cal}}, y_{\text{cal}}, z_{\text{cal}})/M]_{\text{Measured}}}{[D(A_{\text{cal}}; x_{\text{cal}}, y_{\text{cal}}, z_{\text{cal}})/\Psi_0]_{\text{Calculated}}}(1 + b(A_{\text{cal}})). \tag{6}$$

Following equations (4) and (5), one can easily identify the models needed for implementation of the formalism. The presence of f and Ψ_{hsc}/Ψ_0 in equation (4) tells us to model the primary energy fluence and head scatter fluence prior to running the dose calculation engine $d(\ldots)$. Collimator backscatter to the monitors requires a model of its own as indicated by $b(A)$. The strength of the formalism is that the required models are exchangeable, i.e. it does not matter if an analytical, kernel or Monte Carlo based model is used to execute the role of $d(\ldots)$ as long as it provides dose per incident energy fluence.

4.2. Dose-to-collision kerma formalisms

An early concept for beam calibration was exposure, which formed the basis for formulation of the tissue-air-ratio method of dose calculation. Through the work of Bjärngard and others, a kerma based formalism has been developed. Displacement effects from charged particle transport, although small except in the build-up region, can be included in order to generalize the approach to be used for higher energies (Hannallah *et al* 1996) but are neglected here for simplicity. Hence, the total dose per primary collision kerma is expressed as

$$\frac{D}{K_c}(A; x, y, z) = 1 + \mathrm{SPR}(A; x, y, z) \tag{7}$$

where SPR is the scatter to primary dose ratio (cf table 1). Collision kerma is proportional to energy fluence through $K_c = (\mu_{en}/\rho)\Psi$, so kerma distributions scales with energy fluence and can be calculated through simple application of attenuation and the inverse square law. Attenuation calculations require a detailed knowledge of the spectrum or careful experimental analysis in narrow beam geometries (cf Bjärngard and Shackford 1994, Bjärngard *et al* 1989, 1990, Bjärngard and Vadash 1995, Karlsson *et al* 1993). The dose per monitor unit follows as

$$\frac{D}{M} = \left[\frac{K_c(A_{cal}; x_{cal}, y_{cal}, z_{cal})}{M}\right]_{\text{Measured}} \left[\frac{K_c(A; x, y, z)}{K_c(A_{cal}; x_{cal}, y_{cal}, z_{cal})}\right]_{\text{Calculated}} \times S_c(A)(1 + \mathrm{SPR}(A; x, y, z)). \tag{8}$$

In this formalism lateral head scatter variations are neglected through the use of position-independent output factors in air. The evaluation for arbitrary fields can be done by various methods and is not dependent on the dose formalism.

4.3. Empirical dose-to-dose ratio formalisms

The aim of empirical dose-to-dose ratio formalisms is to arrive at the dose per monitor unit by using as few and standardized measurements as possible. This is achieved by varying the independent variables one by one and deriving the factor by which the measured dose value changes. In this way, dose is factored into a set of measurable factors ending with a relation to the calibration geometry:

$$\frac{D}{M}(\text{case A}) = \frac{D/M_{\text{case A}}}{D/M_{\text{case B}}} \frac{D/M_{\text{case B}}}{\cdots} \cdots \frac{\cdots}{D/M_{\text{calib.geom.}}} \frac{D}{M}(\text{calib. geom.}). \tag{9}$$

A recent systematic reformulation of the dose-to-dose ratio formalism has been given by Dutreix *et al* (1997). The strength of the formalism lies in that the calculations are simple once the data are available. However, calculations of full spatial distributions are not adequately addressed by this formalism since the factors are rarely spatially separable, i.e.

$$D(x, y, z) \neq D(x_0, y_0, z_0) \frac{D(x)}{D(x_0)} \frac{D(y)}{D(y_0)} \frac{D(z)}{D(z_0)} \tag{10}$$

but this is of minor importance since the formalism is mainly intended for manual spot checks of monitor units calculated by treatment planning systems. The major limitation is that, in principle, complex treatments require separate measurements for each beam set-up.

4.4. Renormalizations

The basic difference between the reviewed formalisms is that different models with different sets of elementary data, such as pencil kernels or tissue phantom ratios, generate dose normalized to different entities. Renormalization of the calculated dose to a common

calibration condition provides a link between different formalisms. As an example, in the implementation of equation (5), the ratio Ψ_0/M_0 is not explicitly required since the formalism may equally well be viewed as a simple renormalization:

$$\frac{D}{M}(A; x, y, z) = \left[\frac{D(A_{\text{cal}}; x_{\text{cal}}, y_{\text{cal}}, z_{\text{cal}})}{M}\right]_{\text{Measured}} \times \left[\frac{d(A : x, y, z)}{d(A_{\text{cal}}; x_{\text{cal}}, y_{\text{cal}}, z_{\text{cal}})}\frac{1+b(A_{\text{cal}})}{1+b(A)}\right]_{\text{Calculated}}. \tag{11}$$

Equation (9) already contains the measured dose value for the calibration field and the renormalization is obvious.

5. Heterogeneity corrections and scatter dose calculation based on broad beam data

The classical approach for dose calculations to a heterogeneous geometry is to correct the dose acquired for a similar but homogeneous geometry. More recent methods calculate the dose directly by including effects from heterogeneities directly into the models. The situation for modelling of volume scattering effects is similar; a number of methods estimate scatter variations while newer models make calculations based on particle or energy transport directly. Despite the use of full simulation models in modern treatment planning systems, simple approaches are still needed and will most likely be further refined for independent checks of the treatment planning results. There are numerous reviews on inhomogeneity correction methods in the literature, often categorized according to their ability to regard anatomical information (Purdy and Prasad 1983), their way of modelling primary and scattered photon dose components (Wong and Purdy 1987), or the amount of CT density information they use for 3D scattered photon dose calculations (Mohan *et al* 1981, Purdy *et al* 1987, Wong and Purdy 1990, 1992, Bloch and Altschuler 1995). Here various methods for estimation of heterogeneity corrections will be briefly summarized according to the dimensionality of the density information these apply (Purdy 1992). A section on miscellaneous techniques for scatter dose estimation is also included.

5.1. Corrections and scalings for heterogeneities

It is often practical to describe the influence of a tissue heterogeneity as a perturbation of the dose to a homogeneous phantom exposed to an identical irradiation. Commonly, a correction factor is defined from the dose (dose rate) ratio measured for the heterogeneous geometry versus the homogenous, cf equation (9)

$$\left[\frac{D}{M}(\ldots)\right]_{\text{Heterogeneous object}} = \text{CF}(\ldots)\left[\frac{D}{M}(\ldots)\right]_{\text{Homogeneous water phantom}}. \tag{12}$$

Most methods to estimate the heterogeneity correction factor are based on a direct raytrace from the primary source to the point of interest. More elaborate methods such as the ETAR method also exist.

5.1.1. One-dimensional heterogeneity correction of broad beam dose. Methods that use densities only along primary photon paths and hence approximate the patient as a stack of semi-infinite slabs, different for each ray, are here classified as 1D. This type of correction is also widely applied in 3D treatment planning systems where density information is commonly derived from CT images. As the correction is rather independent of the methods used to arrive at the dose for the homogeneous case, a broad variety, too numerous to be reviewed here, of

combinations of methods have been used to calculate the dose. It must be emphasized that even if all these combinations of methods generate a full 3D dose distribution they cannot be considered a correction for heterogeneities in three dimensions.

5.1.1.1. Primary beam effective pathlength methods (EPL). The idea of EPL methods is to scale the broad beam dose distribution by the factor that the primary energy fluence at the depth of calculation has actually changed as compared with the homogeneous case. Assuming water-like media, the density averaged depth at geometrical depth z is given by

$$z' = \frac{1}{\rho_w}\int_0^z \rho(z'')\,dz'' \tag{13}$$

where ρ_w is the density for water and the 'local' density $\rho(z'')$ is (in most cases) estimated from CT images. In applications, z' either replaces the depth variable directly or is used to construct a correction factor. The following four examples show, in order, the effective attenuation method (where μ_w is usually estimated from PDD data), the ratio of TAR, the effective SSD method (Cunningham 1972, Purdy and Prasad 1983), and the isodose shift method (Greene and Stewart 1965, Sundblom 1965), respectively:

$$\begin{aligned}
\mathrm{CF}(z) &= e^{-\mu_w(z'-z)} \\
&= \mathrm{TAR}(A, z')/\mathrm{TAR}(A, z) \\
&= \frac{\mathrm{PDD}(A, z', \mathrm{SSD})}{\mathrm{PDD}(A, z, \mathrm{SSD})}\left(\frac{\mathrm{SSD}+z'}{\mathrm{SSD}+z}\right)^2 \\
&= \frac{\mathrm{PDD}(A, z-n(z-z'), \mathrm{SDD})}{\mathrm{PDD}(A, z, \mathrm{SDD})}
\end{aligned} \tag{14}$$

where n is an empirical constant. EPL methods model the primary dose variation satisfactorily, except for situations of severe charge particle disequilibrium such as for higher beam energies in the lung. However, the amount of scattered radiation reaching the calculation point depends on both position of the inhomogeneity as well as on its size. Therefore, when calculating dose far away from an inhomogeneity, EPL methods give results with acceptable errors but for a complex heterogeneous medium and for dose calculations within or in the near vicinity of an inhomogeneity, EPL methods yield large errors (Sontag and Cunningham 1977).

5.1.1.2. Power-law (Batho) method. This method was suggested by Batho (1964) as an *empirical* correction to account for both primary beam attenuation and scatter changes within water and below a single slab of lung material with density relative to water of 0.35. Sontag and Cunningham (1977) generalized the method to handle arbitrary densities and non-water-like materials. Later, Webb and Fox (1980) and Cassell *et al* (1981) went further to allow for multiple regions of slab-like materials. Finally, El-Khatib and Battista (1984) and Thomas (1991) showed that the correction factor should be based on build-up depth-shifted TMRs instead of the initially proposed TARs such that

$$\mathrm{CF}(z) = \frac{(\mu_{en}/\rho)_N}{(\mu_{en}/\rho)_w}\prod_{m=1}^{N}(\mathrm{TMR}(z - z_m + z_{bu}))^{(\mu_m-\mu_{m-1})/\mu_w} \tag{15}$$

where μ_m and μ_w are the linear attenuation coefficients of the material in layer m and water respectively, $(\mu_{en}/\rho)_N$ is the mass energy absorption coefficient of the material in layer N, z_{bu} is the buildup depth and z_m is the distance along the beam from the surface to the layer m in the phantom. Wong and Henkelman (1982) have demonstrated the fundamental limitations of the original and generalized Batho method through a theoretical analysis carried out on

 8267

the calculation of primary and first scatter photon dose and supported their findings with experimental verifications. According to these, the generalized power-law method provides an acceptable approximation below a single inhomogeneous layer with an extent larger than that of the field size and electron density less than that of tissue. In the extreme situation of a large inhomogeneity with electron density greater than tissue and large field sizes, the method has been proved to be inadequate with errors up to 10%. If the electron density (relative to water) of the inhomogeneous layer is greater than unity, the power-law method overestimates the dose. For the calculation of dose within an inhomogeneous layer, generalized corrections perform well when the relative electron density of the inhomogeneity is less than unity, but results become progressively worse for denser layers and larger field sizes. It is also clear that the method is limited by the requirement of lateral charged particle equilibrium, i.e. small fields of high-energy beams in lung may cause very large errors (El-Khatib and Battista 1984, Thomas 1991).

5.1.1.3. Corrections based on 1D convolutions with exponentials. Several authors have constructed heterogeneity corrections based on essentially the following statement: 'particles released from primary interactions in a (thin) slab deposit their energy proportional to exponential attenuation functions of the radiological distance from that slab'. Convolving the attenuated deposition with the sources of primary interactions Iwasaki (1985, 1990), Petti *et al* (1987) and van de Geijn (1987) demonstrated results for a photon beam entering a phantom of media layered across the beam direction. Lateral effects were commonly included by using data dependent on field size, typically scaled from water data using O'Connor's theorem. In the work by van de Geijn *et al* , it was shown that their method is accurate within 1–2% for all energies examined regardless of the field size. At 10 MV and higher energies the performance on the central axis was considered quite acceptable with errors increasing with decreasing field sizes. More recently, Ahnesjö *et al* (1992b) have derived a simple formula for correction of dose due to scattered photons (equation (21) in their paper). Combined with an effective pathlength scaling of the primary dose, their method proved to yield a total dose accuracy on the 2–4% level except for situations of lateral charged particle disequilibrium (small fields of high-energy beams in low-density regions) where larger errors occurred.

5.1.2. The equivalent tissue air ratio method (ETAR). The equivalent tissue air ratio method (ETAR) (Sontag and Cunningham 1978a, b) was developed as a method to be directly implemented on computers available at the time of its introduction. Despite its limitations, ETAR was widely implemented in commercial treatment planning systems during the 1980s and is still in use in many clinics. It can be considered as the first practical method for computerized treatment planning using CT data. Using the density scaling theorem (O'Connor 1957), the TAR in a field of radius A_r at depth z in a medium of density ρ relative to water, is equal to $\mathrm{TAR}(\rho A_r; \rho z)$; the tissue air ratio in a unit density medium for field size ρA_r and depth ρz. The ETAR correction factor is formally defined as

$$\mathrm{CF} = \mathrm{TAR}(\rho A_r; \rho z)/\mathrm{TAR}(A_r; z) \tag{16}$$

which is strictly valid only for homogeneous, non-unit-density water-like media. The application to heterogeneous geometries is carried out by replacing ρA_r by $\tilde{\rho}_r A_r$ and ρz by $\tilde{\rho}_z z$ where $\tilde{\rho}_r$ and $\tilde{\rho}_z$ are 'effective' densities, estimated by 'practical', approximate algorithms. The method devised by Sontag and Cunningham (1978b) derives $\tilde{\rho}_z$ by averaging CT values along primary photon ray paths (similar to the effective pathlength methods). For $\tilde{\rho}_r$ they stated that there must exist a set of weights such that the mean density weighted over the entire irradiated volume equals $\tilde{\rho}_r$. Such weights should consider the conditions of irradiation, the

irradiated medium and the location of the calculation point, a procedure that in itself would require extensive modelling. An approximate estimation of weighting factors was proposed in the appendix of the original paper (Sontag and Cunningham 1978b): all CT slices were 'coalesced' to form an effective slice at the plane of calculation; thus reducing the 3D volume summation to a 2D summation since the primarily goal at the time was to calculate dose to a single slice. The approximation is not easily interpretable for error estimates. In any case, the ETAR method represented a major improvement over the state of the art at the time of its introduction and pioneered the use of CT distributions in treatment planning.

Due to its widespread implementations several workers have attempted to improve the performance of ETAR in various aspects. Woo *et al* (1990) aimed to improve modelling of different dose categories by using data from Monte Carlo simulations and pointed out analogies of their extended concept with the kernel methods. Redpath and Thwaites (1991) generalized the original concepts to account for beam modulation and derive 3D dose distributions. Yu and Wong (1993) recasted the basic formulations and applied the convolution theorem to design a calculation approach orders of magnitude faster than the original ETAR.

5.2. Scatter dose estimation

Calculation of the scatter dose as a function of field size and shape is a long-standing issue in photon beam calculations. Two common approaches are reviewed here, one that circumvents explicit integrations by using data measured for simple square or circular fields of 'equivalent' sizes, and one that employs parametrized scatter dose representations to facilitate integrations over the field shape.

5.2.1. Equivalent field sizes. The simplest way to estimate the total dose at a point in a homogeneously irradiated phantom from a beam with non-standard (rectangular or irregular) shape is to use measured data for an 'equivalent field', i.e. a square or circular field of such size that gives the same scatter dose as the non-standard field (Johns *et al* 1949, Day 1950, 1978). Based on a linearization of the scatter contribution, Bjärngard and Siddon (1982) proved that the radius of an equivalent circular field for a square field with side s is given by

$$r = 2s \ln(1 + \sqrt{2})/\pi = 0.5611s \tag{17}$$

which explained the earlier work of Day (1972, 1978) who semiempirically used this relation and derived tables of equivalent fields (independent of energy and depth) that are still successfully used (BJR 1996). Investigations using measured data (Tatcher and Bjärngard 1993) and Monte Carlo calculated pencil beam kernels (Ahnesjö *et al* 1992b) have further confirmed the work by Bjärngard and Siddon and supported the use of the method as a convenient estimate of depth dose distributions in simple geometries. In an analogous manner, the equivalent square s of a rectangular field of dimensions $L \times W$ is derived from

$$s = \frac{2LW}{L+W} Y(L/W) \tag{18}$$

where Y is defined as the elongation factor. Values of Y are tabulated by Bjärngard and Siddon (1982). Equation (18) without the elongation correction factor has been known as the area-to-perimeter ratio (4A/P method) and was first used empirically by Sterling *et al* (1964) and later examined by others (Patomaki 1968, Worthley 1966, Wrede 1972). Essentially, the 4A/P method equates the central axis dose of any field with that of a circular field of the same area, which can lead to serious errors (Day and Aird 1996).

5.2.2. Scatter dose representation and integration techniques. The representation of scatter dose and related integration techniques has been addressed by many authors. Such scatter functions can be derived from measured data, analytical calculations or Monte Carlo calculations. The most well known scatter representation is scatter air ratios (SAR) and differential scatter air ratios (dSAR) as defined by Cunningham (1972) (cf section 2.4). Recent work by Sontag and Ray (1995) introduces the calculation of dSAR (from measured data and analytically) that correspond to different orders of photon scattering. An alternative representation of scatter dose is given using scatter phantom ratios and in particular scatter primary ratios (SPR). Bjärngard and Petti (1988) and later Bjärngard and Vadash (1995) have showed semiempirically that SPR at depth z on the central axis (excluding the buildup region) of homogeneous fields with square side s could be approximated by

$$\mathrm{SPR} = \frac{asz}{ws + z} \tag{19}$$

where a and w are estimated from a limited set of measurements. Furthermore, Bjärngard and Vadash (1995) showed that a is related to the probability that scattered photons are produced, and w to their directional distribution. Storchi and van Gasteren (1996) and Sätherberg *et al* (1996) have also developed parametrizations, as functions of beam quality index, for the scatter to primary dose ratios at 10 cm depth in water phantoms. The parametric representations of scatter data, combined with equivalent field size approaches have proven to be of great value for quality control of both measured data and monitor unit calculations from treatment planning systems (Bjärngard *et al* 1997). Furthermore, pencil kernel data can be derived from radial differentiation of scatter functions thus forming a bridge between traditional empirical methods and modern Monte Carlo methods for kernel determination (Ceberg *et al* 1996).

Dose for general field shapes can be derived through integration over the field aperture of appropriately differentiated scatter data or representations such as equation (19) (cf section 7.3 on pencil kernel methods). An early integration method that has been widely applied is the approximate summation technique according to Clarkson (1941) and implemented into a widespread computer program by Cunningham *et al* (1972). Here the field around the point of calculation is separated into a number of angular segments and the scatter contribution from each segment is estimated by use of measured data. The Clarkson method works well for simple field shapes but runs into methodological problems for complicated field shapes. Siddon *et al* (1985) developed a far more general method based on triangular decomposition of the surface integral. This method works for any field shape that can be described by a set of closed polygons (where blocks are described by polygons of opposite direction) and can also utlilize parametrizations such as equation (19).

6. Implicit modelling of scattered particle transport

Implicit modelling of particle transport through scaling operations is less computationally intense than a full, explicit modelling of particle interaction and transport. Hence, implicit modelling methods have been extensively studied as illustrated in the following with three different approaches, namely the differential scatter air ratio, the delta volume method and the kernel based methods. The first two approaches have never been widely applied but deserve attention since they highlight some aspects common to the modern kernel based models. Kernel based methods, finally, are treated separately in section 7.

6.1. The differential scatter air ratio model (DSAR)

The first method to address the 3D problem of dose to heterogeneous phantoms by scaling first- and higher-order scatter as first scatter was the differential scatter air ratio method as proposed by Beaudoin (1968). Cunningham (1972) later published how scatter air ratios (SAR) can be derived from tissue air ratios (TAR), and further how the former can be differentiated to give differential scatter air ratios (dSAR). These describe contributions to dose at a point in water from photons scattered in surrounding volume elements as a function of the distance to that point. Scatter dose contributions $(\mathrm{dSAR}/\mathrm{d}V)_{\mathrm{medium}}$ at point r in an inhomogeneous medium from a volume element at r' are expressed in the DSAR method as

$$\left(\frac{\mathrm{dSAR}}{\mathrm{d}V}\right)_{\mathrm{medium}} = \left(\frac{\mathrm{dSAR}}{\mathrm{d}V}\right)_{\mathrm{water}} \rho_e(r') f_1(r') f_2(r, r') \tag{20}$$

where $\rho_e(r')$ is the electron density relative to water at the scattering site, f_1 is a factor describing the attenuation of the beam relative to water between source and volume element ΔV and f_2 is a factor describing the attenuation of secondary photon fluence relative to water along the path between ΔV and the dose calculation point. Factors f_1 and f_2 can be derived from the Klein–Nishina cross sections and the relative electron density at the dose calculation point and along the assumed rectilinear path of scatter transport respectively. The approach is differentiated enough to model beam modifiers and irregular fields and an accuracy of $\pm 2\%$ in simple heterogeneous geometries has been reported (Cunningham and Beaudoin 1973, Larson and Prasad 1978).

Although DSAR methods employ 3D scatter ray-trace procedures and measured SAR to represent the overall scattering strength of a voxel, it has been shown to be inaccurate when modelling the irradiation of a heterogeneous phantom with large field sizes and at low energies (Sontag 1979). This has been interpreted as that a first scatter ray-trace model is incompatible with the use of SAR data which implicitly contain contributions from multiply scattered photons (Cunningham 1972). The original DSAR model has been examined only for ^{60}Co beam. Implementation at higher energies would suffer from the lack of electron transport modelling due to difficulties in representing primary and scattered photon contributions from measured (extrapolated) zero area TAR and SAR respectively. Redpath (1995) described a simplified implementation of the DSAR philosophy where electron transport is ignored and scatter dose is calculated in an approximate manner by assigning factor f_2 of equation (20) equal to the relative electron density at the point of interaction (similarly to simplified FFT kernel based convolution methods; cf section 7.2.2.2). Another limitation of the DSAR method is that backscatter is not modelled due to the difficulty in deriving explicit backscatter differential scatter air ratios from TAR (Wong and Purdy 1990).

6.2. The delta volume model

Although some earlier work might have been labelled as the delta volume method due to the use of differential SAR from small volume elements, it is the work by Wong and Henkelman which is generally recognized as the delta volume method (Wong and Henkelman 1983, Wong *et al* 1984). Dose at a point in a heterogeneous medium is calculated as a sum of the primary dose, an augmented first-scatter dose component and an approximate residual multiple- scatter component. Relative primary dose is obtained similarly to the DSAR method from the knowledge of the primary intensity in air and the density along the path of the primary photons. The augmented first-scatter component includes the part of the second-order scatter that was considered to be effectively transported as first scatter (scattering angles less than 45°). Both these components were pre-calculated

as a kernel using cross-sectional data and scaled to actual geometries by explicit ray tracing. The residual multiple-scatter dose component is modelled in a way that resembles ETAR scaling of SAR with density. In this term, empirical data representing the dose perturbation from introduction of a small void in an otherwise homogeneous water phantom are used.

The physics behind the scatter modelling in the delta volume method has been well examined and justified for ^{60}Co beams through investigations of the scatter dose properties (Wong *et al* 1981a, b). The method succeeds in satisfying the two constraints identified by Wong and Henkelman (to correctly calculate the dose to (a) water with a small void and (b) homogeneous non-water medium), it approximately calculates the multiple-scatter dose component, and since it is using augmented scatter values and multiple-scatter perturbation values from each scattering element it directly accounts for backscatter. The computational burden, absence of electron transport modelling, reliance on experimentally cumbersome data and lack of development for higher beam energies have made the method of less interest for implementation into commercial treatment planning systems.

7. Kernel based models (convolution/superposition)

Kernel based convolution/superposition models are a family of models with roots in the imaging world. Analogous to image formation, the dose deposition is viewed as a superposition of appropriately weighted responses (kernels) to point irradiations. Under conditions where the kernels are spatially invariant, the superpositions can be efficiently evaluated by means of convolutions. The kernels, representing the energy transport and dose deposition of secondary particles stemming from a point irradiation, are not usually accessible through measurements but are very simple to calculate by use of Monte Carlo particle transport codes. The earliest record of point kernels in dosimetry known to the authors is by Loevinger (1950, 1956). The buildup region from a betatron photon beam depth dose was explained by a one-dimensional forward scatter function approach by Johns *et al* (1949). Later Roesch (1958) introduced kerma (which he called 'kerm') and defined an 'influence function' that distributes the kerma energy at a point into absorbed dose at surrounding points. Dutreix *et al* (1965) used buildup curves for various narrow circular fields to determine the approximate shape of the Roesch influence function. Brahme (1977) used an equivalent concept to calculate restricted mass energy absorption coefficients for use in dosimetry. Dean (1980) used point kernels for 1.25 MeV gamma rays together with experimental data from LiF thermoluminescent dosimeters, for the calculation of the relative amount of scatter dose. Schoknecht (1971) and Ulmer (1982) used pencil kernels in a convolution process to demonstrate calculations of dose distributions.

The potential for kernel based models in treatment planning did not attract much interest until 1984 when the concept was brought forward by several independent investigators (Ahnesjö 1984, Boyer and Mok 1984, Chui and Mohan1984, Mackie and Scrimger 1984) and later worked out in more detail (Boyer and Mok 1985, Mackie *et al* 1985, Mohan *et al* 1986, Ahnesjö *et al* 1987). Although the formulation of the method in its basic form is simple and appealing, the demands on computer time combined with the need for modelling of various second-order beam characteristic effects have delayed its clinical implementation until recently. Kernel models can explicitly handle the degree of freedom posed by modern treatment machines without any major approximations and it is therefore generally anticipated that they will be the workhorse for conformal therapy applications (Webb 1993). In the following sections we shall review the kernel superposition approach more in detail.

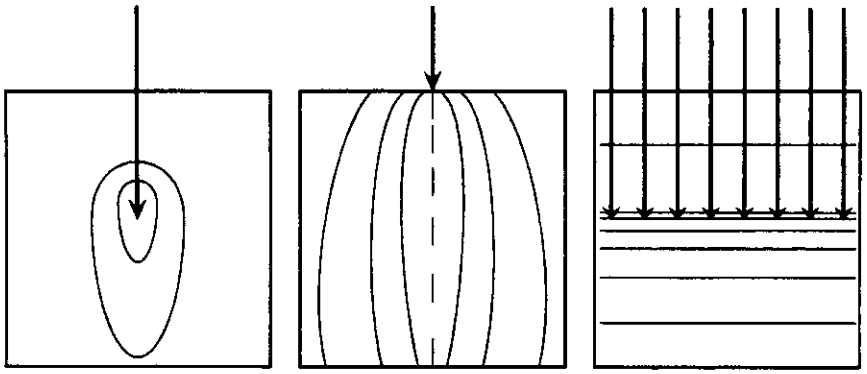


Figure 6. Irradiation geometries for point kernels (left), pencil kernels (centre) and planar kernels (right). Isodose curves are shown as full curves.

7.1. Energy deposition kernels

Photons can travel large distances unaffected and the energy and direction of a primary photon is therefore independent of where it interacts. The energy deposition by secondary particles around a primary photon interaction site is, in homogeneous media, independent of the location of the site and can be described by a kernel. Energy deposition kernels (EDK) are defined as the distribution of energy imparted to volume elements (per unit volume) in a medium, commonly water, due to an elemental photon beam incident at the origin of coordinates of the kernel. Energy deposition kernels are categorized according to the geometry of the elemental beam that delivers the incident energy. Essentially three different kernel geometries have been defined (see figure 6). The kernel describing the pattern of energy deposited in an infinite medium around a primary photon interaction site is known as a *point kernel* (called dose spread array by Mackie *et al* (1985), differential pencil beam by Mohan *et al* (1986) and point spread function by Ahnesjö *et al* (1987)). A *pencil kernel* describes the energy deposition in a semi-infinite medium from a point monodirectional beam and a *planar kernel* describes the forward and backward energy spread from primary interactions located in a plane, laterally oriented in an infinite broad beam. Sometimes also a fourth type, a *rotated* kernel, describing the deposition of energy due to convergent irradiation of a rotationally symmetrical phantom, has been used for inverse calculations (see section 10.1).

There are several possibilities for normalization of kernels, dependent on the formulation of the dose equation they will be part of. The common approach is to normalize to the radiant energy to be distributed by the kernel, i.e.

$$h(r) = \frac{\mathrm{d}\varepsilon}{R\,\mathrm{d}V} \Rightarrow \iiint_{\infty} h(r)\,\mathrm{d}V \equiv 1 \tag{21}$$

where $\mathrm{d}\varepsilon$ is the mean energy imparted in the volume element $\mathrm{d}V$ due to the interactions the radiant energy R undergoes before it is deposited as dose. Alternative approaches have been used by Boyer (1988) who normalized the kernel towards the fluence of the primary particles, and by Mohan *et al* (1986) who normalized the kernel to the number of interacting primary particles. Kernels are often separated into different dose categories according to the scattering history of the depositing particle (cf figure 1). Separating a point kernel into two parts, one for

the primary dose (h_p), and one for the phantom scatter dose (h_s), the kernel integral is closely related to the energy absorption coefficients as pointed out by Mackie *et al* (1988) and Boyer (1988). Using the normalization in equation (21), it follows that

$$\iiint_{\infty} h_p(r)\,dV = \frac{\mu_{en}}{\mu} \qquad \iiint_{\infty} h_s(r)\,dV = \frac{\mu - \mu_{en}}{\mu} \tag{22}$$

where μ_{en} is the linear energy absorption coefficient and μ the linear attenuation coefficient. This comprises a useful check for verifying the generation of primary and scatter kernels and is also of importance for constructing corrections for beam quality variations.

The complexity of coupled electron/positron–photon transport limits the possibilities for analytical methods for calculating kernels. The standard method for calculation of kernels is therefore direct Monte Carlo simulations of the particle transport (Ahnesjö *et al* 1987, Mackie *et al* 1988, Mohan *et al* 1986). This is straightforward, although somewhat demanding on CPU time. Analytical *modelling* of the primary dose kernel has been done by Wang *et al* (1995) using Gaussian multiple scatter theory. In photon transport, the first scatter kerma kernel is particularly simple to derive and has been utilized in convolution schemes proposed by Boyer and Mok (1985) and by Nilsson and Knöös (1992). Both groups also devised approximate methods for derivation of the residual, multiple scatter kernels. Boyer and Mok modelled the kernel for multiple scattered photons assuming that they are isotropically distributed around a first scatter interaction site. Nilsson and Knöös also proceeded from the first scatter kernel and used a buildup factor to model the multiple scatter contribution. Analytical *fitting* to simplify the use of kernels has been done for point kernels (Ahnesjö 1989, Ahnesjö and Mackie 1987) and for pencil kernels by Ahnesjö *et al* (1992b). The approach used by Ahnesjö for polyenergetic point kernels was to model the kernel as mediated by rectilinearly transported particles with exponential attenuation and inverse square divergence according to

$$h(r) = \frac{A_\theta e^{-a_\theta} + B_\theta e^{-b_\theta}}{r^2} \tag{23}$$

where A_θ, a_θ, B_θ and b_θ are fitting parameters depending on the scattering angle θ. The separation of primary and scatter was such that the first term mainly describes the primary and the second term the scatter dose fraction.

Energy deposition distributions must, when calculated by Monte Carlo, be scored in finite voxels. The larger the voxels the better the statistical accuracy and therefore all workers have utilized the rotational symmetry of spherical or cylindrical binning to increase the scoring volume of distant voxels. Applications of kernels are facilitated through binning these in a problem-specific coordinate system. For example, Fourier transform convolution requires a Cartesian binning, pencil kernels are most natural to use in cylindrical coordinates, and radial scaling of point kernels for heterogeneities is best carried out using the radial bins of a spherical coordinate system. Scaling and change of the coordinate system requires rebinning or interpolation between kernel values scored in neighbouring bins. In low-gradient regions this is trivial, but care must be taken close to the primary interaction site since there is a singularity at the origin, cf the denominator of equation (23). Although the point kernel value is infinite at the origin, integrals over finite volumes are always finite. The physical interpretation of the singularity is best understood in terms of particle track density of primary released electrons. Since they all stem from one point, the track density becomes infinite and so does the dose, given constant ionization per track length. Also, photon pencil kernels are singular along the symmetry axis since the track density becomes infinite with all primary electron tracks originating from the axis. Special methods for kernel rebinning based on an energy conservative

volume overlap technique have been proposed by Eklöf *et al* (1990). For Cartesian- cylindrical conversions, Rathee *et al* (1993) used interpolating functions renormalized within each original bin to improve the rebinned result.

Direct experimental validation of point kernels is not possible because one cannot force photon interactions to take place at a single location in an extended absorber. However, the reciprocity principle can be applied to design experiments as carried out by O'Connor and Malone (1989). Recently, pencil beam dose kernels were determined for 6 and 25 MV x-rays by fitting analytical models to quantities derived from broad beam data (Ceberg *et al* 1996). Iwasaki derived *forward and backward spread functions* from measured data in water to calculate primary and scatter dose separately (Iwasaki 1985, Iwasaki and Ishito 1984).

Energy deposition kernels are invaluable tools for understanding qualitative aspects of dose distributions. At low energies (< 1 MeV), the electron range is very much shorter than the photon mean free path. A considerable portion of the primary photon energy is also further transported to yield scatter dose as shown in figure 7. At very high energies such as 50 MeV the electron track lengths are of the same order as the photon mean free paths and only a minor part of the primary photon energy is transferred to scatter dose as illustrated in figure 7.

7.2. Point kernel models

The calculation of dose from point kernels can be described as a two-step procedure as sketched in figure 8. In the first step the energy released in the patient through attenuation of the primary photons is calculated by ray-tracing primary photon trajectories, including beam modulators, etc. The raytrace is normally performed in a Cartesian matrix (cf Siddon 1985) with interaction data mapped from CT scans to represent the patient. In the second step, dose is calculated by superposition of appropriately weighted kernels. Following the common kernel normalizations in equation (21), the dose equation for monoenergetic irradiation of a homogeneous phantom with a parallel beam follows as

$$D(\boldsymbol{r}) = \iiint_V T(\boldsymbol{s})h(\boldsymbol{r}-\boldsymbol{s})\,\mathrm{d}^3\boldsymbol{s} \tag{24}$$

where $T(\boldsymbol{s})$ is the *terma* (*t*otal *e*nergy *r*eleased per *ma*ss (Ahnesjö *et al* 1987)) from the primary photon energy fluence $\Psi(\boldsymbol{s})$ in the volume element $\mathrm{d}^3\boldsymbol{s}$. The integration variable $\boldsymbol{s}$ in equation (24) is terma oriented and the purpose of the kernel is to weight the energy transfer from all irradiated $\boldsymbol{s}$ to $\boldsymbol{r}$.

Through variable substitution, equation (24) is equivalent to

$$D(\boldsymbol{r}) = \iiint_V T(\boldsymbol{r}-\boldsymbol{s})h(\boldsymbol{s})\,\mathrm{d}^3\boldsymbol{s} \tag{25}$$

in which the integration variable $\boldsymbol{s}$ is kernel oriented and the kernel weights the energy transfer from all $\boldsymbol{r}-\boldsymbol{s}$ to $\boldsymbol{s}$. The reciprocity between photon interaction and dose deposition sites has been discussed by Hoban *et al* (1994), cf section 2.3. When equations (24) or (25) are computer coded as discrete summations to yield dose distributions as a function of $\boldsymbol{r}$, one has to loop over the locations of both $\boldsymbol{s}$ and $\boldsymbol{r}$. Choosing the loops over the $\boldsymbol{r}$ locations to be outermost yields dose values point by point. If instead the outermost loops are chosen to be over the $\boldsymbol{s}$ locations the dose distribution is gradually built up and no point dose values are ready until the full distribution is ready. Mackie *et al* (1985) labelled these different loop orders after

 1075

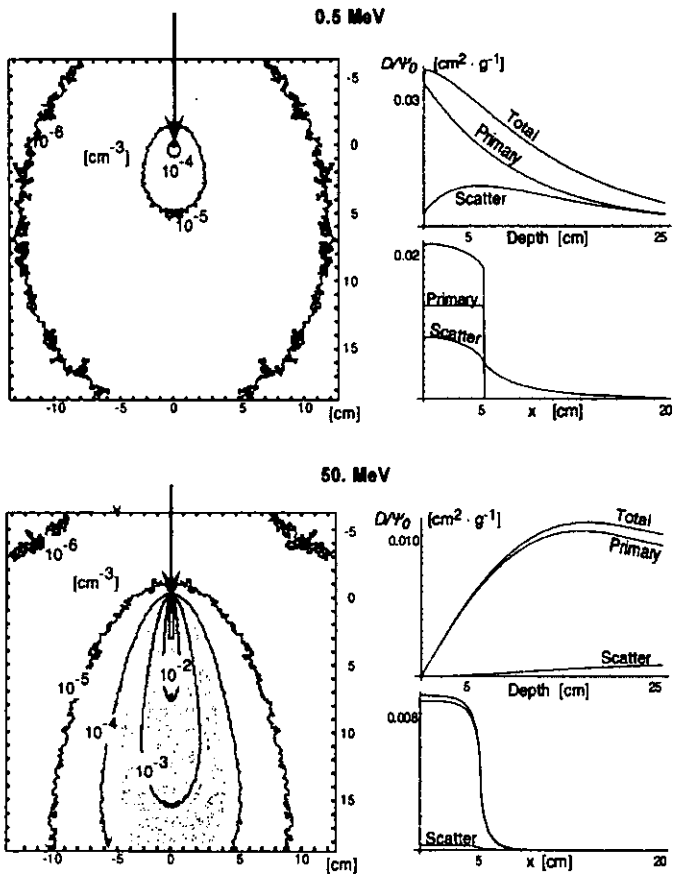


Figure 7. Point kernels in isolevel format (left) and depth doses and profiles (right) for a 10×10 cm^2 field at infinite SSD. The grey area of the 50 MeV point kernel shows the lateral extension of the primary dose. The corresponding area at 0.5 MeV is so small that it is not resolved (from Ahnesjö 1992).

the outermost loops as the '*dose deposition point of view*' and the '*interaction point of view*' respectively.

The dose equation (24) is exactly valid for an arbitrary distribution of the fluence of monoenergetic photons incident in a parallel beam on an infinite medium—a highly idealized situation. We will in the following sections discuss the adaptation of point kernel methods to more general conditions.

7.2.1. Basic generalizations and approximations. As reviewed by Battista and Sharpe (1992), the considerations that impose approximations to point kernel superposition models are (a) the spectral and geometrical properties of clinical x-ray sources, (b) the heterogeneous medium of finite extent (a patient!) and (c) the time constraints imposed by interactive treatment planning.

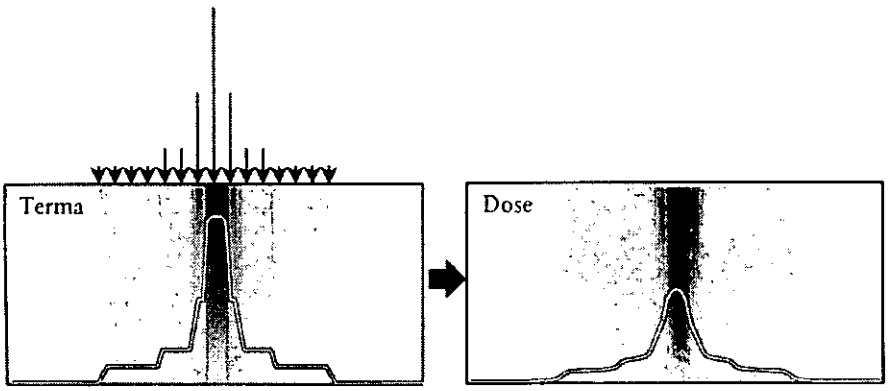


Figure 8. The initial step (left) of point kernel models is to ray trace the incident beam for calculation of the energy transferred to secondary particles, i.e. the terma. In a subsequent step the energy of the secondary particles are transported and deposited as dose (right). Due to the energy transport by secondary particles, dose patterns are always smoother than the corresponding patterns of primary beam energy release.

7.2.1.1. Generalization for primary beam spectral variations. Polyenergetic beam sources (cf section 9.1.1) can be considered by generalizing equation (24) to

$$D(\boldsymbol{r}) = \int_E \iiint_V T_E(\boldsymbol{s}) h(E, \boldsymbol{r} - \boldsymbol{s})\, \mathrm{d}^3 s\, \mathrm{d}E \tag{26}$$

where the energy dependency is included by using an energy-dependent kernel and terma differential in energy. Here, the terma differential in energy is given by

$$T_E(\boldsymbol{r}) = \frac{\mu}{\rho}(E, \boldsymbol{r}) \Psi_E(\boldsymbol{r}) \tag{27}$$

where $\mu/\rho(E, \boldsymbol{r})$ is the mass attenuation coefficient of the primary photons of energy E and $\Psi_E(\boldsymbol{r})$ the energy fluence, differential in energy, of primary photons at $\boldsymbol{r}$. Terma calculation and subsequent superposition over an energy dimension require repeated spatial integrations (see equation (26)) and is computationally expensive. Representing the energy spectrum with a single (mean) energy does not result in accurate depth dose values. Boyer *et al* (1989) found that five energy bins were enough to represent the spectrum for a 6 MV Siemens machine. Five bins were also used by Zhu and Van Dyk (1995) who investigated the sensitivity in depth dose from variations in each spectral bin. Hence, a straightforward discrete implementation of equation (26) requires that the spatial convolution must be repeated at least five times with considerable timing drawbacks. The spectrum variations to consider are (a) 'depth hardening' of the primary beam due to filtration in the irradiated object and (b) 'off-axis softening', i.e. lateral variation of the spectrum due to decreasing beam hardening in the flattening filter with off-axis distance, cf Lee (1997). Several approaches have been proposed to circumvent the need for explicit calculations for each energy interval.

For depth hardening, Metcalfe *et al* (1989, 1990) concluded that terma is the major factor in determining the shape of the calculated dose distribution. Therefore, in polyenergetic beams, it is important to include spectral influence on the terma distribution. Papanikolaou *et al* (1993) carried out a full calculation for polyenergetic terma and showed that a polyenergetic kernel should be averaged using terma-weighted contributions for the spectral bins. Beam hardening

effects on the polyenergetic kernel were accounted for by applying a precalculated depth-dependent correction factor derived from the ratio of depth dose curves obtained from a full calculation (equation (26)) and the single polyenergetic kernel superposition. Liu *et al* (1997c) replaced the correction factor by use of a kernel, interpolated for each depth from polyenergetic terma-weighted kernels pre-calculated for the surface and two more depths. However, attenuation coefficients, and consequently also terma distributions and kernels (in the normalization according to equations (21) and (22)), are only weakly dependent on energy (Ahnesjö 1987, Ahnesjö *et al* 1987). Hence, Hoban *et al* (1994) and Hoban (1995) proved that the separation of the energy diffusion process into a primary dose convolution and a scatter dose convolution, suggested by Ahnesjö (1991), yields very accurate results. In essence, they made the following approximation

$$D(\boldsymbol{r}) = \int_E \iiint_V T_E(\boldsymbol{s}) h(E, \boldsymbol{r}-\boldsymbol{s})\,\mathrm{d}^3\boldsymbol{s}\,\mathrm{d}E$$
$$\approx \iiint_V P(\boldsymbol{s})\tilde{h}_\mathrm{p}(\boldsymbol{r}-\boldsymbol{s})\,\mathrm{d}^3\boldsymbol{s} + \iiint_V S(\boldsymbol{s})\tilde{h}_\mathrm{s}(\boldsymbol{r}-\boldsymbol{s})\,\mathrm{d}^3\boldsymbol{s} \tag{28}$$

where the released energy distributions for primary, P (i.e. the collision kerma), and scatter, S, are given by

$$P(\boldsymbol{r}) = \int T_E(\boldsymbol{r})\frac{\mu_\mathrm{en}}{\mu}(E)\,\mathrm{d}E \tag{29}$$

$$S(\boldsymbol{r}) = \int T_E(\boldsymbol{r})\left(1 - \frac{\mu_\mathrm{en}}{\mu}(E)\right)\mathrm{d}E \tag{30}$$

with the corresponding kernels (cf equation (22)) weighted by the terma at a certain depth z_0 and renormalized through

$$\tilde{h}_\mathrm{p}(\boldsymbol{r}) = \frac{\int \Psi_E(z_0)\mu(E)h_\mathrm{p}(E,\boldsymbol{r})\,\mathrm{d}E}{\int \Psi_E(z_0)\mu_\mathrm{en}(E)\,\mathrm{d}E} \tag{31}$$

$$\tilde{h}_\mathrm{s}(\boldsymbol{r}) = \frac{\int \Psi_E(z_0)\mu(E)h_\mathrm{s}(E,\boldsymbol{r})\,\mathrm{d}E}{\int \Psi_E(z_0)(\mu(E)-\mu_\mathrm{en}(E))\,\mathrm{d}E} \tag{32}$$

to yield unity integrals over infinite space. The separation enables the study of primary and scatter dose separately and also provides the means for straightforward implementation of off-axis softening corrections (Saxner and Ahnesjö 1998). The kernel definition depth z_0 used by Hoban *et al* was set to the phantom surface, but an arbitrary depth such as 10 cm can be used with insignificant difference in calculated dose. Hoban (1995) also showed that beam hardening translates to an (almost linear) increase of the mean primary photon energy with depth and the consequence of this is shown to be a linear increase of the ratio of collision kerma to terma with depth. Terma calculation time can be a significant part of the overall dose calculation time (Reckwerdt and Mackie 1992) and the linearity can be exploited to simplify calculations of the P and S distributions respectively.

7.2.1.2. Beam divergence. For divergent beams, there is an inverse square factor reduction of the primary photon fluence with depth, a linear increase of each field dimension with depth and a rotation, 'tilting', of the kernels (see figure 9). Papanikolaou *et al* (1993) investigated the approximation of applying the inverse square correction outside of the inner summation loop to increase computation speed, i.e. the inverse square correction is applied at the dose deposition site instead of the primary interaction site. It was found

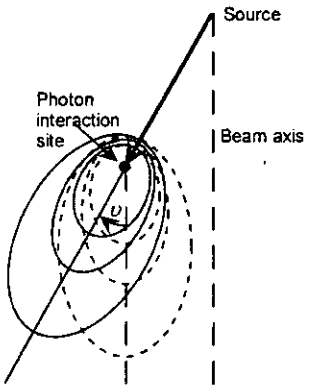


Figure 9. Kernels should ideally be 'tilted' by an angle ν versus the beam axis to align with the primary photon direction at the interaction site.

that approximation compensates for not aligning kernels to the fan-like geometry of the beam. Many implementations neglect the effect of kernel tilting and a study on the validity of parallel kernel approximation was carried out by Sharpe and Battista (1993). They found that in combinations of extreme cases such as small SSD, large field size and high energy, errors above 3% are likely to be observed. Locally in penumbras larger errors are generated as artefacts. Nevertheless, they concluded that using parallel kernels was an acceptable approximation for most clinical situations. Liu *et al* (1997c) made detailed comparisons of different approaches to kernel tilting and pointed out that the computation of the tilt angle ν is most efficiently calculated in the 'dose deposition point of view' since a complete coordinate transform between the kernel system and the beam system can be avoided.

7.2.1.3. Tissue heterogeneity density scaling and finite patient extent. The transfer of energy by first scatter photons depends on the constitution of the medium between the primary photon interaction site and the dose deposition point. The deposition of energy mediated by multiply scattered particles depends on the medium located elsewhere, but it depends more on the medium close to the initial direction then on media far off, since the scattering cross sections are largest in the forward direction. This justifies the common approach of scaling all dose fractions of a point kernel h_{ρ_0}, calculated for a homogeneous medium of mass density ρ_0, by the mean electron density between the point s of energy release and the point r of energy deposition, i.e.

$$h_{\text{het}}(s, r) = \frac{\rho(r)}{\rho_0} c^2 h_{\rho_0}[c(r - s)] \tag{33}$$

where

$$c = c(s, r) = \int_0^1 \rho_{\text{rel}}[s - \ell(s - r)]\, \mathrm{d}\ell \tag{34}$$

in which ρ_{rel} is the relative number of electrons per volume as compared with the reference medium. The convolution integral in equation (24) is then replaced by the superposition

 279

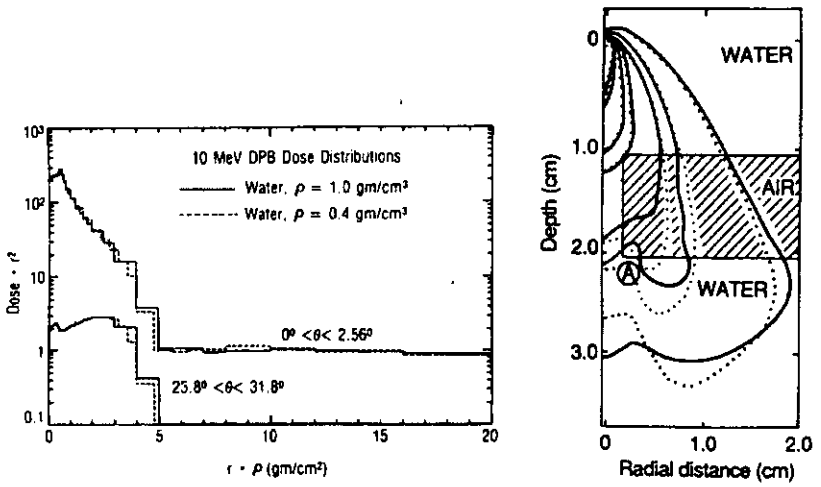


Figure 10. Left: Comparison of density scaled point kernels in homogeneous media with kernels generated directly with Monte Carlo. Data are presented along radial lines out from the interaction centre and multiplied by the radius squared. (From Mohan *et al* (1986), with permission.) Right: Comparison in isolevel format of a density scaled point kernel (dotted) and a Monte Carlo kernel (full) in a heterogeneous geometry with an air ring. The largest deviation was found to be at point A. (From Woo and Cunningham (1990), with permission.)

integral

$$D(r) = \iiint_V T(s)\frac{\rho(s)}{\rho_0}c^2 h_{\rho_0}[c(r-s)]\,\mathrm{d}^3 s. \tag{35}$$

This scaling has been evaluated by Mohan *et al* (1986) (see figure 10) and is consistent with the theorems of O'Connor and Fano. The density scaling is used to avoid the tedious task of calculating an 'exact' kernel for every situation. The problem has also been investigated by Woo and Cunningham (1990) who compared Monte Carlo kernels for a sample of inhomogeneous geometries with density scaled *water* kernels. They concluded that although the deviations between the kernels could locally be substantial, as shown in figure 10, the overall effect on dose was less severe due to averaging from surrounding kernels. Density scaling of a discrete kernel is numerically the same type of problem as the rebinning of a homogeneous kernel into bins of sizes as scaled according to the inhomogeneities. This requires special care in high-gradient regions, cf discussion on rebinning in section 7.1. The effect of high kernel gradients is probably the main reason for differences in results from apparently equivalent methods of kernels as reported by Wong and Purdy (1990).

For the primary dose kernels, two groups have worked towards improving the density scaling procedures by employing the Fermi–Eyges electron scattering theory. Keall and Hoban (1995) extended the rectilinear density scaling method by multiplying c in equation (36) with a factor based on the Fermi–Eyges calculated lateral planar fluence distribution. Although this extra scaling was implemented in the direction of the beam axis only, the results improved as shown in figure 11, but with an increase in calculation time by a factor of 3. Yu *et al* (1995) proposed to transport the electrons separately, by an electron pencil beam model in which the medium is approximated as layered along the transport direction. This was also shown to model well the effect on dose due to the presence of materials with different atomic

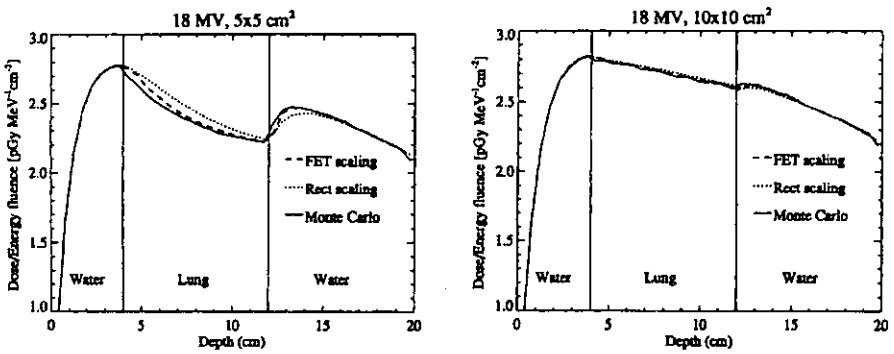


Figure 11. Depth dose curves for a 5×5 cm^2 beam (left) and a 10×10 cm^2 beam (right), for 18 MV in a water–lung–water phantom. Results of Monte Carlo (full curves), rectilinear density scaling of point kernels (dotted curves), and electron fluence scaling based on Fermi–Eyges theory (dashed curves) are shown. (From Keall and Hoban (1995), with permission.) The error from density scaling of the primary dose vanishes when the electrons contributing dose to a point originate within the same density regions.

numbers. Sauer (1994, 1995) suggested a method that considered the difference in stopping power and range between water and a high atomic number (Z) material, where the energy deposited in the high-Z material is calculated by rectilinear path scaling using the mean mass stopping power ratio of the high-Z material to water and by multiplying the energy deposition value with the mass stopping power ratio at the site. In addition, to account for the increased lateral diffusion of electrons in the high-Z material, the energy deposited was adjusted by averaging with energy deposited at adjacent angular positions. Recently, Keall and Hoban (1996a) proposed a method where they circumvented the kernel scaling problem by explicitly transporting electrons using pre-generated Monte Carlo electron track data. This required prolonged calculation times, a factor of 30 as compared with a straightforward kernel superposition.

The dose deposited through multiply scattered photons was considered by Boyer and Mok (1985) to be rather diffused and they employed a diffusion theory for modelling dose from these photons. Similarly, Mackie *et al* (1985) chose, according to the average density, an appropriate scatter kernel from a database of kernels generated in water-like media of different densities. Kernel representation of the dose from multiply scattered photons is also affected by the common approximation of using a very large, or infinite, phantom while generating the kernel. This will result in an overestimation of the dose near media boundaries, such as in cases of tangential irradiation, due to photons considered to be backscattered through the surfaces facing empty space as if it was water everywhere. Superposition calculations for a 4 MV beam in a 'half phantom' geometry (without surface curvature) have demonstrated errors of several per cent within the first 3 cm from the phantom edges (Aspradakis 1996). At higher energies, however, the kernel distributions are more forward directed and dose due to multiply scattered photons is less and the error decreases. A detailed investigation on photon beam exit dose was carried out for ^{60}Co and 24 MV beams by Woo (1994). He confirmed the findings of Mohan *et al* (1986), namely that a convolution calculation using an invariant kernel underestimates the dose drop at the beam exit surface due to a reduction of backscatter not being accounted for.

7.2.2. Calculation algorithms. Despite the continuous rapid development of computers with ever- increasing computational speed, straightforward application of kernel superposition is still time consuming enough to impede its full implementation into clinical routine. Several projects have demonstrated speed improvement by use of specialized hardware (cf Murray *et al* 1991, McGary and Boyer 1997). Although photon transport has an inherent parallelism, the increasing system complexity and lack of standardization has so far limited the use of specialized parallel architectures in commercial treatment planning systems. Hence, there has been an intense search for effective algorithms to speed up calculations, especially since the demand for dose computations in treatment planning is foreseen to increase from a more widespread use of iterative optimization methods.

7.2.2.1. Direct summation. Several methods can be used for direct summation of density scaled kernels in dose calculations. Straightforward discrete implementation of equation (24) in a homogeneous medium requires N^6 multiplications and summations to calculate the result to N^3points. The inclusion of kernel scaling as in equation (35) will further increase the number of operations to N^7. Direct summation is therefore very time consuming, and several techniques and approximations have been explored in order to increase the calculation speed. These include the use of Fourier transforms, the collapsed cone approximation and correction factor interpolation.

7.2.2.2. Fast transform techniques. The use of fast transform convolution techniques, such as the fast Fourier transform (FFT), to carry out the discrete convolution (Ahnesjö 1984, Boyer and Mok 1984, 1985, Murray *et al* 1989) scales the number of operations for the homogeneous case essentially with the 3D Fourier transform requiring $3N^2$ 1D transforms each of $N \log_2 N$ operations, i.e. the approach scales as $N^3 \log_2 N$ instead of the direct summation proportionality N^6. Transform based fast convolutions require a spatially invariant kernel and several attempts have been made to circumvent the invariance requirement. Boyer and Mok (1986) proposed a method that scaled the scatter dose kernel by the density at the scatter site only. This method has been tested (Wong and Purdy 1990, Zhu and Boyer 1990) and found to yield encouraging results in a number of situations. The method has been improved (Ahnesjö 1987, Wong *et al* 1996) to include the densities at both the scattering and the receiving sites for scaling of the scatter kernel. However, all these methods fail to scale the primary dose due to high kernel gradients. The problem of lateral electron transport in FFT convolution has been studied further by Wong *et al* (1997) who presented an improvement on the calculation of primary dose along the central axis in cases of electronic disequilibrium and in the penumbra region by scaling field sizes at each depth with local effective densities. These are computed, in a manner resembling that used in the ETAR model, by convolving the density at the interaction site with the primary kernel for water.

7.2.2.3. Collapsed cone convolution. The collapsed cone method proposed by Ahnesjö (1989) applies an angular discretization of the kernel which, together with the parametrization in equation (23), enables an efficient approach for energy transport and deposition. Angular discretization of a parametrized point kernel yields, for each discrete angular sector (cone) Ω_i, the energy deposition per radial distance as

$$\iint_{\Omega_i} \frac{h_{\rho_0}}{\rho}(r, \Omega) r^2 \, \mathrm{d}^2\Omega = A_{\Omega_i} \mathrm{e}^{-a_{\Omega_i} r} + B_{\Omega_i} \mathrm{e}^{-b_{\Omega_i} r}. \tag{36}$$

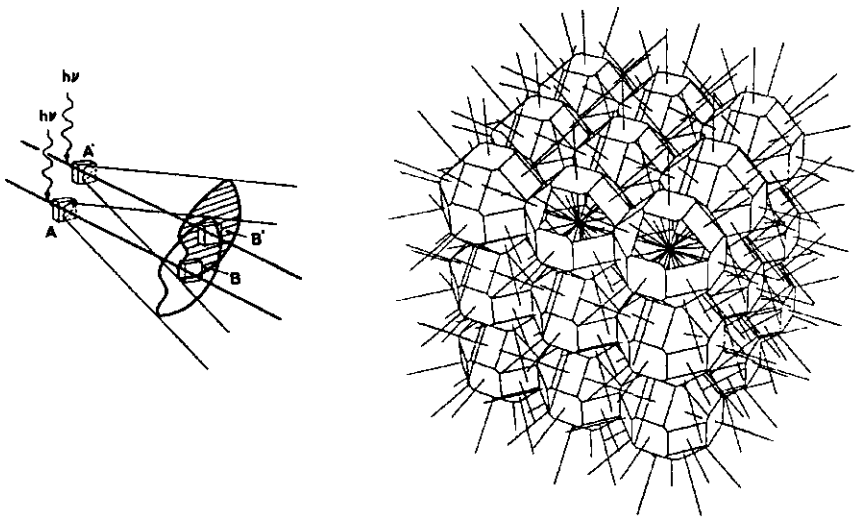


Figure 12. Left: A consequence of the collapsed cone approximation to transport the energy along a discrete axis is that the energy that should have been deposited in voxel B′, from interactions at the vertex of the lower cone, is deposited in voxel B and vice versa. This displacement grows with distance; however, the first scatter fraction decreases with increasing distance, making the approach acceptable as the total energy deposited is conserved. Most energy is deposited close to where it is released, making displacement errors less important since it takes place mainly within voxels, as shown by the voxels A and A′. (From Ahnesjö 1989.) Right: Example of a simple lattice of cone axes made to cover the 3 × 3 × 3 calculation voxels with the 26 discrete cone directions used in this case. (From Ahnesjö 1991.) In this example all transport lines intersect at each voxel centre but this is not required (Reckwerdt and Mackie 1992) as long as each voxel is intersected by each direction.

Notice that the inverse square of the radius cancels due to the increasing cross section of the cone Ω_i with increasing radius. When the angular discretized kernel is convolved with the terma distribution, all energy released into the cone direction Ω_i from volume elements on the cone axis is approximated to be rectilinearly transported, attenuated and deposited in volume elements on that axis, i.e. the cones are *collapsed* onto their axes (see figure 12). A lattice of transport lines, representing cone axes, is constructed to cover the irradiated volume such that each direction intersects every calculation voxel. This requires a parallel subset of lines for each discrete direction of the 'collapsed' kernel, which can be arranged in several ways (Ahnesjö 1989, 1991, 1997, Reckwerdt and Mackie 1992) (see figure 12 for a simple example). Due to the exponential description of the kernel, the energy transport along a line can be expressed analytically resulting in recursive formulae that only need to be evaluated once for each voxel on that line. Kernel scaling for the heterogeneities is performed during the recursion, both for the primary and scatter dose kernels (i.e. both terms in equations (23) and (36)). The recursions pass each voxel at least once per direction. When each point is calculated individually, the number of operations will be proportional to MN^4, where M is the number of conical sectors (angular bins). If instead the dose is calculated in one sequence for a bulk of N^3 points the total number of operations needed in heterogeneous media is proportional to MN^3. In a similar method Reckwerdt and Mackie (1992) bypassed the exponential parametrization by using the result of an initial ray-trace along each transport line to look-up the radiological distance between each pair of release/deposition points. This

 5283

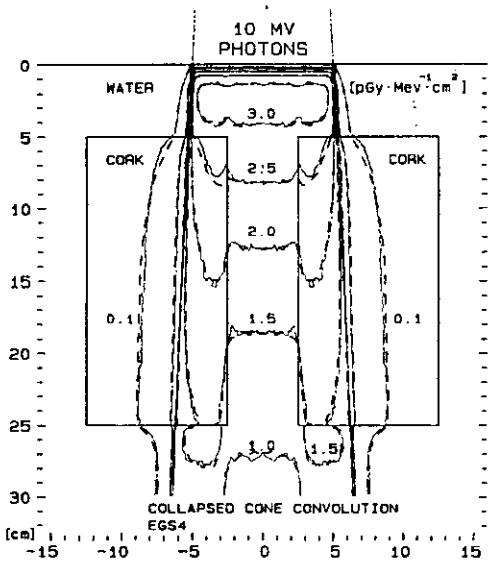


Figure 13. Dose from a 10×10 cm^2 10 MV beam in a lung phantom where the hatched isodoses show the result from collapsed cone calculations and solid isodoses the result from Monte Carlo. Note the penumbra broadening in lung and buildup after lung which is well reproduced using the collapsed cone approximation. (From Ahnesjö 1989.)

results in an MN^4 algorithm since for each point the other points on the same line have to be visited separately.

The collapsed cone convolution method has been verified against Monte Carlo generated dose distributions for slab and mediastinum-like phantom geometries (Ahnesjö 1989; see figure 13), and for clinical beams in homogeneous media (Lydon 1998), in both cases demonstrating generally very good results.

7.2.2.4. Correction factor interpolation. Aspradakis and Redpath (1997) devised a technique for reducing computation time in a superposition model. The method is based on the assumption that the conventional inhomogeneity correction method adequately predicts the shape of the dose distribution over a limited distance. Hence correction factors, defined as the ratio of dose values from superposition to those from the conventional (broad beam, fast, inhomogeneity correction) algorithm calculated on a coarse matrix and further interpolated onto a fine grid are applied to relative dose values calculated on the fine grid by the fast conventional method. This method requires the generation of two 3D matrices; one from the conventional model on a fine matrix and the other from superposition on a coarse matrix. The speed enhancement as compared with a full calculation approach depends on the resolution ratio. As an example, Aspradakis and Redpath demonstrated a 128-fold improvement in speed without significant loss in accuracy when the correction factor was carried out every 4.0 cm along the direction of the beam and every 2.0 cm across the beam in comparison to a grid separations of 0.5 cm in all dimensions. However, in regions of high dose gradients, a finer resolution might be required.

 3284

7.3. Pencil kernel models

A pencil kernel describes the energy deposited in a semi-infinite medium from a point monodirectional beam (see figure 6). For the purpose of treatment optimization, Gustafsson *et al* (1994) used a very general formulation of the radiotherapy dose calculation problem:

$$D(\boldsymbol{r}) = \iint_{s} \int_{E} \iint_{\Omega} \sum_{m} \Psi^{m}_{E,\Omega}(\boldsymbol{s}) \frac{p^{m}}{\rho}(E, \Omega, \boldsymbol{s}, \boldsymbol{r})\, \mathrm{d}^2\Omega\, \mathrm{d}E\, \mathrm{d}^2\boldsymbol{s} \tag{37}$$

where $\Psi^{m}_{E,\Omega}(\boldsymbol{s})$ is the energy fluence differential in energy E and direction Ω for beam modality m and $(p^m/\rho)(E, \Omega, \boldsymbol{s}, \boldsymbol{r})$ is the corresponding pencil kernel for energy deposition per unit mass at $\boldsymbol{r}$ due to primary particles entering the patient at $\boldsymbol{s}$. For routine dose calculations the approach of equation (37) is still impractical since the pencil kernel is variant and in principle unique for each combination of $\boldsymbol{s}$ and $\boldsymbol{r}$. In equation (37) the beam that yields this kernel is of infinitesimal cross section whereas in practice Gustafsson *et al* discretized the kernels to describe the energy deposition from divergent, finite width beam elements (often referred to as bixels) which converted the integral to a matrix multiplication form. The discrete pencil kernel problem has been studied in more detail by Bourland and Chaney (1992) and Ostapiak *et al* (1997). Their study was motivated by the potential savings in computation time for superposition models since use of pencil kernel models can be viewed as a 'pre-convolution' of point kernels over the depth dimension. The pencil kernels are commonly determined using Monte Carlo methods, either directly or as a superposition of point kernels. Monte Carlo methods have been used by a number of workers (Mohan and Chui 1987, Ahnesjö and Trepp 1991, Ahnesjö *et al* 1992b, Bourland and Chaney 1992). Pencil kernels have also been estimated experimentally using scatter factor differentiation (Ceberg *et al* 1996, Storchi and Woudstra 1996).

The integration over the field aperture is greatly simplified by use of spatially invariant kernels and various techniques have been developed to numerically evaluate such integrals. When the lateral dose distribution at a particular depth is of interest, standard convolution methods using transforms can be applied (Mohan and Chui 1987, Ahnesjö and Trepp 1991). Bortfeld *et al* (1993) showed that the dose at arbitrary points can be interpolated from two profiles, essentially representing primary and scatter dose, using a depth dependent weighting technique derived from a singular value decomposition of the kernel data. Storchi and Woudstra (1996) combined a pencil kernel approach by computing a sparse lattice of points using transform technique with an interpolation along beam lines to reconstruct dose at arbitrary points. Ahnesjö *et al* (1992b) noted that pencil kernels, expressed in cylindrical coordinates (r, z), can be accurately parametrized as

$$\frac{p}{\rho}(r, z) = \frac{A_z \mathrm{e}^{-a_z r} + B_z \mathrm{e}^{-b_z r}}{r} \tag{38}$$

where A_z, a_z, B_z and b_z are functions of depth. They also developed a numerical integration technique based on a combination of the parametrization with a triangular decomposition of the field shape (similar to that of Siddon *et al* (1985)), and fluence weighting to allow for both intensity modulation and complicated field shapes.

Pencil kernel models are effectively hybrid algorithms that fully account for beam modulations and field shapes but rely on broad beam scaling/correction methods to handle heterogeneities and patient outlines (cf section 5.1.1). Provided there are properly normalized fluence and kernels, the dose is calculated in absolute units that can be used to derive output factors. The two main accuracy limitations on pencil kernel models are for heterogeneities (Knöös *et al* 1995) and for scatter dose calculations in phantom sizes that deviate substantially from the size for which the pencil kernel is determined (Hurkmans *et al* 1995). Nevertheless, the

inherent flexibility to model lateral fluence variations combined with computational efficiency has made the pencil kernel method popular, in particular for use in optimization of intensity modulation (Bortfeld *et al* 1994, Gustafsson *et al* 1994, Söderström *et al* 1993, Chui *et al* 1994).

8. Explicit modelling of scattered particle transport

8.1. Random methods, Monte Carlo

The widespread use of Monte Carlo applications in medical physics has been reviewed by Mackie (1990) and Andreo (1991). Photons have a limited number of interactions before they are absorbed, which makes it easy to simulate all interactions directly. In electron transport, however, the number of Coulomb interactions with atomic nuclei is so large that a direct Monte Carlo simulation is impractical. This has motivated the development of 'condensed' history techniques where different 'microscopic' interactions are classified into groups to provide a detailed 'macroscopic' representation of the particle transport (Berger 1963). The condensing assumes the medium to be locally homogeneous, making it non-trivial to study interface effects in detail. The grouping may be implemented in different ways using different approximations (Andreo and Brahme 1984, Bielajew and Rogers 1987, Kawrakow and Bielajew 1998). Dose scoring for treatment planning is, however, not as critical as simulation of ionization chamber responses. Hence, Monte Carlo is considered to be the primary method for bench-marking of faster dose calculation approaches.

For photon beam dose calculation, the long mean free paths between interactions means that the energy is transferred to the medium in local fractions over large volumes and necessitates simulation of a very large number of photon histories to reduce statistical uncertainty. The secondary electrons released by the photons smooth the dose distribution, which is more efficient at higher energies due to longer electron tracks. Hence, Monte Carlo is more efficient at higher photon energies than at lower energies making brachytherapy sources the most demanding candidate for direct use of Monte Carlo dose calculations. Long calculation times are a pronounced problem in conformal therapy where optimization of dose distributions using iterative algorithms requires the dose to be recomputed many times during the planning procedure. Hence, it is likely that for another decade photon Monte Carlo will be used for beam characterization, benchmarking and other special studies rather than routine treatment planning. Nevertheless, a very ambitious Monte Carlo treatment planning project has been launched at the Lawrence Livermore Laboratory aiming to model all beam modalities, including photons, by means of stand-alone machines networked to the treatment planning computer (Hartmann Siantar *et al* 1997). Based on the EGS4 code, Wang *et al* (1998) proposed to improve efficiency through the use of variance reduction techniques and density thresholding, the latter to speed up the ray-tracing calculations. For electron beams, several projects report clinical or near-clinical implementations for direct use in treatment planning (Manfredotti *et al* 1990, Neuenschwander *et al* 1995, Keall and Hoban 1996b, Kawrakow *et al* 1996), and it is likely that Monte Carlo calculations will be the clinical workhorse for electron beam calculations in the near future.

8.2. Deterministic methods for particle transport

The most general description of a radiation field at a given time is the phase space density, i.e. the number of particles per six-dimensional 'volume' made up of the three spatial coordinates $\boldsymbol{r} = (x, y, z)$, the particle energy E and the direction $\Omega = (\theta, \phi)$. Since the particles of a

 3286

radiation beam actually move, the phase space can be completely represented by the vectorial energy fluence differential in energy and direction for photons, $\Psi^{ph}_{E,\Omega}$, and for charged particles (electrons and positrons), $\Psi^{cp}_{E,\Omega}$. A number of quantities of dosimetric interest can be derived directly from these fluences and in particular the dose follows according to Rossi and Roesch (1962)

$$D(r) = -\frac{1}{\rho(r)}(\nabla \cdot \Psi^{ph}_{E,\Omega}(r) + \nabla \cdot \Psi^{cp}_{E,\Omega}(r) - q(r)) \tag{39}$$

where ρ is the mass density and q the rest mass changes. From a given initial radiation field, continuity laws, as expressed in the well known Boltzmann equation, govern the resulting fluence distributions. The scattering and absorption processes make the Boltzmann equation difficult to solve except for very simplified cases. Bellman and co-workers (see Bellman and Wing (1975) for an introduction) have made theoretical studies proposing the use of an invariant embedding technique to solve the radiation transport problem in slab geometries. A more general method to numerically solve the Boltzmann equation is the method of discrete ordinates which discretizes the spatial dimensions into voxels, the directions into 'discrete ordinates' (as for the collapsed cone method; cf section 7.2.2.3) and the energy range into 'multigroups' (see Bell and Glasstone (1970) for a general introduction). The discrete ordinates, or S_N, method has long been used for neutron transport calculations in nuclear engineering, and code packages such as DANTSYS are generally available (Alcouffe *et al* 1995). The method has not attracted much interest in the medical physics community and there are only a limited number of papers available including those by Shapiro *et al* (1976) who investigated ^{252}Cf as a brachytherapy source, Uwamino *et al* (1986) who calculated neutron leakage from a medical electron accelerator, Nigg *et al* (1991)who investigated dose distribution analysis in boron neutron capture therapy and Gokhale *et al* (1994) who focussed on inverse methods for beam direction selection in 2D planning. The method is only suited to problems where the particles are subject to a limited number of interactions before coming to rest since the number of iterations is proportional to the number of interactions. Hence, to completely solve the dose deposition in photon beams numerically by purely deterministic methods the electron transport must be solved by a more suitable method such as the phase space evolution originally developed for dose calculation in electron beams (Huizenga and Storchi 1989, McLellan *et al* 1992, Janssen *et al* 1997). The complexity and computational burden for such a complete deterministic approach would probably exceed that of the Monte Carlo approach (Börgers 1998).

9. Beam phase space characterization

Empirical dose calculation models use measured dose data almost directly in calculations. Hence, self-consistency with respect to measured beam data is intrinsic to empirical methods. For more explicit modelling of the radiation energy transport one needs a description of the initial phase space delivered by the beam, i.e. the time-independent energy fluence $\Psi_{E,\Omega}$ differential in energy and direction at all points (x, y, z_0) in a beam reference plane at z_0. However, there are no established methods to determine experimentally such a complete description in the clinical environment. Hence one must use either direct calculations based on the design of the machine or indirect methods based on simple measurable quantities such as dose distributions in a water phantom. In practice it is common to use a combination of both approaches. An excellent review of beam characteristics for both photon and electron beams has been published by Ebert *et al* (1996). The review given here will focus on work applicable to treatment planning with first-principle models.

Monte Carlo simulation of beam transport through clinical treatment heads has proved to be an invaluable tool to determine otherwise inaccessible data. The Monte Carlo approach has recently been greatly facilitated by Rogers *et al* (1995) who introduced the BEAM package tailored to simulate treatment heads by means of the EGS4 code (Nelson *et al* 1985). A similar package has also been designed by Lovelock *et al* (1995). The BEAM package has been used in a series of papers by Liu *et al* (1997a, b, d, 1998) to provide detailed information about phase space distributions for clinical machines. The Monte Carlo system GEANT, developed at CERN, has interfaces to mechanical engineering CAD (computer aided design) systems for handling complex designs. The GEANT system has been utilized by Küster *et al* (1997, 1998) to study multileaf collimators. A shortcoming of Monte Carlo for beam characterization is that the energy of the electron beam hitting the target is often not directly known with the required accuracy. Boyer *et al* (1989) determined the electron beam energy through a calibration of the bending magnet current and found deviations as high as 20% from the specified energy. A practical problem regarding Monte Carlo is that the set up of the simulations requires detailed geometrical knowledge of treatment head details which is not always easily accessible from manufacturers.

First-principle models give the absolute dose per impinging energy fluence. In practice the fluence monitoring is non-trivial since scattered photons from the treatment head add an 'unmonitored' contribution to the fluence, and backscatter into the monitor yields a 'false' contribution to the total signal (cf section 4.1):

$$\frac{\Psi}{M} = \frac{\Psi_0 + \Psi_{\mathrm{hsc}}}{M_0 + M_b}. \tag{40}$$

It is therefore common to describe the unscattered beam and the scattered components of the beam separately, as we will do in the following sections.

9.1. Primary fluence characterization

9.1.1. Beam spectrum and attenuation. Direct Monte Carlo simulations of the treatment head design have been carried out by several workers in order to provide spectral data. Mohan *et al* (1985) calculated, for a series of beam energies, spectra that later were frequently used as a standard set. They also studied the lateral variation in beam spectrum and characterized it in terms of half-value thickness. Lovelock *et al* (1995), Liu *et al* (1997d) and DeMarco *et al* (1998) (who used the MCNP code, (Briesmeister 1988)) applied Monte Carlo techniques to generate beam spectra and they all found that the incident electron beam energy must be tuned in order to get agreement with measured depth dose data.

Deterministic semianalytical calculations model the bremsstrahlung process as an electron diffusion and absorption with simultaneous emission of photons. The most energetic photons are generated in the direction of the electrons. Although the intensity is strongly forward peaked in the initial direction of the electrons, the angular dispersion of the electrons due to multiple scattering makes the unfiltered beam spectrum fairly uniform over the beam aperture. The subsequent filtering to flatten the fluence induces a slight variation such that the peripheral parts are softer due to thinner filters. Examples of works that apply this kind of model are Nordell and Brahme (1984), Ahnesjö and Andreo (1989), Desobry and Boyer (1994), Silver (1994) and Svensson and Brahme (1996). Deterministic models of the bremsstrahlung emission often apply Shiff's formulae as given in the classical review by Koch and Motz (1959). Desobry and Boyer (1991) reviewed the Shiff work and were able to relax some of the approximations that Shiff used.

Reconstructive techniques based on measured depth dose distributions are an appealing approach since the dose calculated using such spectra describes measured data as closely

3288

as possible. To implement such a procedure a database of depth dose distributions from monoenergetic photons calculated by Monte Carlo must be available. The main difficulty in reconstructive techniques from depth dose distributions or attenuation measurements is the poor numerical conditioning of photon spectrum unfolding, which makes the use of different constraints necessary. Furthermore, the dose from charged particle contamination in the buildup region complicates the use of data from that region including the depth of dose maximum. Ahnesjö and Andreo (1989) combined a parametrized model for charged particle contamination with a semianalytical spectrum model whose parameters were varied to minimize the difference between measured depth doses and depth doses reconstructed as the sum of the dose for a pure photon beam and the charged particle dose. In a similar dose reconstructive approach Sauer and Neumann (1990) used general shape properties of realistic spectra expressed as positivity and monotony requirements. Attenuation data are more generally accessible than monoenergetic depth dose distributions and several reconstructive techniques based on attenuation measurements have been proposed (Huang *et al* 1983, Archer *et al* 1985, Piermattei *et al* 1990, Baker *et al* 1995, Baker and Peck 1997, Francois *et al* 1997, Nisbet *et al* 1998). Most of these works also used constraints on the spectral shape to handle numerical conditioning problems.

Methods for direct measurement of photon spectra, although not directly applicable in clinical environments, are of importance for benchmarking other methods. In particular, the data measured by Faddegon *et al* (1990, 1991), who also compared their data with Monte Carlo calculations, have become a standard.

In kernel based dose calculations the spectrum is often assumed to be laterally invariant. This approximation implies that off-axis softening, i.e. the decrease in energy with increasing angle versus the beam axis caused by decreasing flattening filter thickness, is ignored. For point kernel calculations this can be relaxed by considering the effects on the released energy fraction (collision kerma and scatter energy, i.e. equations (29) and (30)) through corrections describing the lateral change in beam penetration. Mohan *et al* (1985) used the angular distributions of photons generated by Monte Carlo techniques to compute off-axis changes in the half-value layer of water. In a broad experimental survey involving 15 different linac beams, Tailor *et al* (1998) showed that the relative change in the narrow beam half-value layer of water of the collated data was well described by a third degree polynomial of the off-axis angle. The earlier data from Yu *et al* (1997) and Bjärngard and Shackford (1994) were close to the data of Tailor *et al*. Although general parametrizations now exist, off-axis beam quality variations depend on the material of the flattening filter (Zefkili *et al* 1994) and should therefore be at least checked as part of the machine commissioning procedure.

9.1.2. Lateral fluence distribution. To achieve high accuracy, dose calculations must model the lateral fluence distribution as delivered by the open beam (not only for open beams but also as input for subsequent modulations of the beam). The common machine design paradigm is to shape the flattening system to deliver a uniform dose at some depth. Phantom scatter dose has its maximum at the beam centre, and the flattening system is designed to compensate for that by increasing the primary fluence with increasing radius (fluence 'horns'). In their Monte Carlo study, Lovelock *et al* (1995) pointed out that beam profiles were more sensitive indicators to variations of the electron beam energy than the depth dose curves. Hence, lateral profiles cannot entirely be predicted by Monte Carlo means and must be determined through measurements. Since detecting fluence is less developed than detecting dose, it is more common to use dose profiles and adapt the fluence profiles in calculations to fit measured dose rather than using measured fluence profiles. Boyer *et al* (1989) and Liu *et al* (1997d) assumed rotational symmetry characterized by a small set of profile measurements. Treuer *et al* (1987)

and Ahnesjö and Trepp (1991) worked out procedures to allow for full lateral mappings of general, non-rotational symmetrical beams. Ahnesjö and Trepp used radial dose scans rotated in 10° to 15° increments to achieve full mapping of the field. Most points are thereby taken in the 'curative region' and drifts in the experimental set-up are cancelled by renormalization to the beam axis point common to all scans.

9.1.3. Beam modulation. There are several methods for modulating the fluence of treatment units, as reviewed by Brahme (1987). The present most common technique is to use wedge shaped absorbers that influence not only modulation but also the beam quality. Hence, wedge factors expressed as fluence ratios will differ from wedge factors expressed as dose ratios. An important consideration in wedge dosimetry is the eventual effects the quality shifts have on different experimental techniques. Weber *et al* (1997) proved that buildup caps for detectors in fluence measurements could be made by any material when used *within* that same beam quality. Heukelom *et al* (1997) noted a difference in results between various cap materials when used for wedge factor determination. For dose calculations, Liu *et al* (1997b) ray-traced several energy bins through both the wedge and the patient. To reduce the number of ray-traces in the patient, Ahnesjö *et al* (1995) treated the quality shifts separately by correcting the separation of the released energies for the primary dose and scatter dose fractions, respectively, suiting the formalism as outlined by equations (28)–(32).

Modulation by dynamic collimation using the blocks, i.e. the dynamic wedge technique (Kijewski *et al* 1978), is relatively straightforward to model since the beam quality is not affected and excellent results have also been demonstrated (Weber *et al* 1996, Liu *et al* 1998, Storchi *et al* 1998). Somewhat more complicated to model is the modulation from multileaf blocks run in either 'step-and-shoot' mode (Bortfeld *et al* 1994, Webb 1998a, b) or in continuous mode (Convery and Rosenbloom 1992, Svensson *et al* 1994) since leaf interference yields spots with aberrant 'tongue-and-groove' underdosage and 'interleaf leakage' overdosage (van Santvoort and Heijmen 1996). At present there is great interest in exploiting the potentials in dynamic collimation motivated by the opportunities seen in treatment optimization utilizing non-uniform beams without increasing the labour cost of delivering the desired modulations.

9.2. Scatter from the treatment head

As stated in section 2, a significant amount of the treatment beam consists of scattered photons and secondary particles from practically all irradiated components of the treatment head. There is a large literature on experimental determination and empirical data fitting of quantities for 'output' characterization, intended for direct practical application. The magnitude and the relative importance of scatter from the various parts of the treatment head, however, have often been matter of confusion. The output factor in air is often called collimator scatter factor, indicating that the collimators are the most scattering parts. This is not the case, and it is now well established that the largest contributions originate in the flattening filter and close to the target in the primary collimator as experimentally showed through the work of Kase and Svensson (1986) and Jaffray *et al* (1993). For open beams on conventional machines, head scattered photons are of the order of one-tenth of the total energy fluence exiting the machine (Zhu and Bjärngard 1995, Weber *et al* 1997), with some additional per cent for wedged beams (Heukelom *et al* 1994, Liu *et al* 1997d). As noted by Weber *et al* (1997), some published output factors in air for high-energy beams do not describe the photon fluence output due to improper filtering of charged particle contamination. A common classification of head scatter sources, as given in the pioneering work of Nilsson and Brahme (1981) and Luxton and Astrahan (1988), is into scatter from the primary collimator (often merged into the flattening

3290

filter component), the flattening filter, collimator backscatter into the monitor and collimator scatter. Nilsson and Brahme used Monte Carlo calculations and found the head scatter sources to be rather limited (<3%) while Luxton and Astrahan determined these experimentally and found them to be very important (<15%). These seemingly conflicting results are explained by the different treatment head configurations used in each work. Nilsson and Brahme used a configuration designed for a maximum beam diameter of 20 cm with a significantly thinner and hence less scattering flattening filter than that of typical clinical machines. Luxton and Astrahan used a CGR Saturne 25 machine which has a conventional flattening filter design but with collimator trimmers very close to the monitor chamber causing an unusually high amount of monitor backscatter. In the following, the treatment head scatter and related calculation models will be reviewed according to their source of origin.

9.2.1. Photon scatter from the flattening filter. The flattening filter forms an extended source distribution, which determines most of the variation of beam output in air due to variation of the field size. The filter is thickest at its centre. This, and the fact that the primary energy fluence at the exit side is approximately homogeneous, makes it natural to assume that the energy fluence of scattered photons is highest at the centre of the flattening filter and decreases towards the periphery. This has been shown experimentally (Jaffray *et al* 1993), analytically (Ahnesjö 1994) and with Monte Carlo simulations (Chaney *et al* 1994, Liu *et al* 1997d). These simple findings together with the relative importance of their magnitude have motivated the development of methods that model the dose from scattered particles by means of an extended source integration over the parts visible from the calculation's point view (Ahnesjö *et al* 1992a, Dunscombe and Nieminen 1992, Ahnesjö 1994, Chaney *et al* 1994, Sharpe *et al* 1995, Liu *et al* 1997d) (see figure 14).

9.2.2. Wedge and compensator scatter. Second to structures closest to the target, hard wedges or compensating filters are the most important scatter sources in clinical beams. Analytical calculation models based on first scatter integration over the scattering device (Ahnesjö *et al* 1995, Islam and Van Dyk 1995), and an 'extended phantom concept' using precalculated modulator kernels superimposed over the modulator within the calculation point of view of (Liu *et al* 1997c), have all shown good results.

9.2.3. Collimator scatter. A model for explicit modelling of scatter from the collimators has been presented by Ahnesjö (1995). The model is based on kernels representing the relative energy fluence of photons scattered off a collimator block per unit block length. The kernels can then be applied to the current field of calculation and integrated around the collimating periphery to yield the total collimator scatter energy fluence at the point of calculation. The results show that in most cases scatter from the movable collimators contributes of the order of approximately 1% to the total dose and hence is relatively insignificant. Exceptions yielding larger contributions were found to be for collimators very close to the source, i.e. the primary collimator, and for geometries involving many layered collimators since each upstream side of a collimator block acts as an inlet for photons to be scattered. For direct Monte Carlo simulations, collimator scatter is a low probability event resulting in poor statistics unless special variance reduction techniques are used. As a consequence of the low amount of collimator scatter in photon beams, the term 'collimator scatter factor' should be avoided in favour of 'output factor in air' or '(treatment) head scatter factor'.

3291

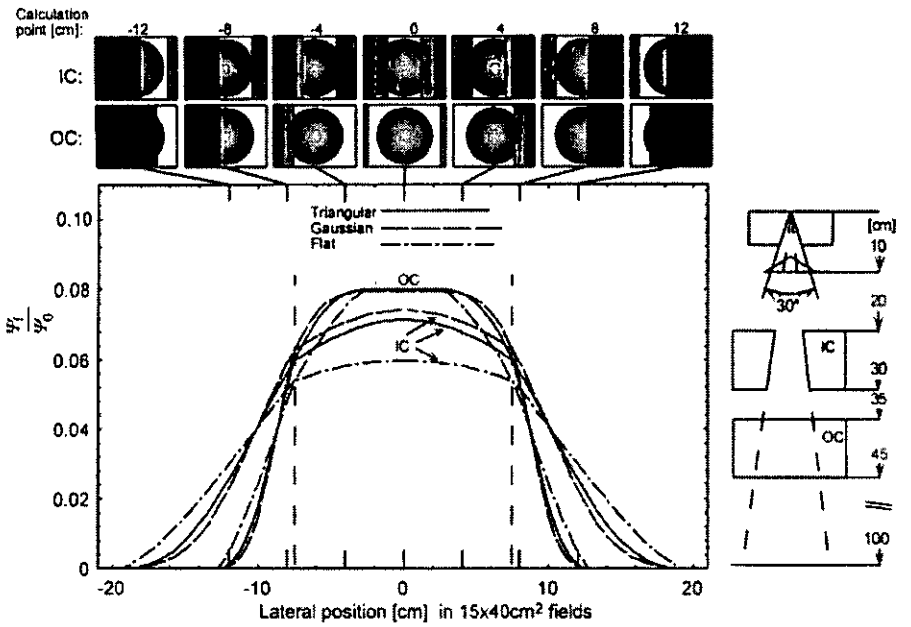


Figure 14. Flattening filter scatter profiles (normalized to the isocentre primary energy fluence Ψ_0) at the isocentre plane for two 15 × 40 cm^2 fields defined by the inner (IC) and outer (OC) collimators. The profiles are along the 15 cm axis; the machine geometry is shown to the right. The calculation point's eye view of the filter at various positions is shown on top of the chart. Three different distributions of scatter release from the filter are compared; a triangular, a Gaussian and a flat (constant) distribution, all normalized to yield 8% scatter at isocentre when the entire filter is viewed. (From Ahnesjö 1994.)

9.2.4. Monitor backscatter. The problem of backscatter to the monitor from the upper side of a collimator has been studied by a variety of methods. Kubo (1989) used a telescopic technique to exclude the scattered components from the readout of an external detector and measured the variation in monitor units delivered per unit external signal. For a Clinac 1800 he found small variations in the order of 1% to 2% but for a Therac 20 machine the backscatter variation was as high as 7.5% between a 2 × 2 cm^2 field and 40 × 40 cm^2 field. Hounsell (1998) also used a telescopic technique and found small variations of the order of less than 1% for an Elekta-Philips SL15 with a protection sheet (3 mm Al) in place between the collimators and the monitor chamber. Hounsell found considerably higher variation when the protection sheet was removed, approximately 5% between a 4 × 4 cm^2 and 40 × 40 cm^2 field. Sharpe *et al* (1995), Yu *et al* (1996) and Liu *et al* (1997b) used the number of linac pulses as independent measure of the primary fluence. Liu *et al* used BEAM/EGS4 to perform Monte Carlo simulations of the phenomenon, and found that the monitor backscatter signal varied between 2% and 5% for the largest and smallest fields with a kapton window monitor chamber. When a protection sheet of aluminium was set in place to stop low-energy charged particles the variation reduced to 0.5–1.0%. Yu *et al* applied the technique to a Varian Clinac 600 C and 2100C and found a variation of approximately 2% for the upper jaws and 1% for the lower pair of jaws at energies above 15 MV and about half those values for 6 MV. Lam *et al* (1998) measured the target charge needed to deliver a given amount of monitor units as a function of collimator setting,

 3292

as it was considered more reliable than the number of linac pulses. On a Varian Clinac 2100C they found a 2.4% variation for the upper jaws and 1.0% variation for the lower pair of jaws.

A calculation model has been developed by Ahnesjö *et al* (1992a) assuming that the backscatter to direct particles signal ratio (cf equations (3)–(6)) can be determined by a proportionality factor k_b times a geometry factor for diffuse radiation, such that

$$\frac{M_b}{M_0} = k_b \frac{z_{\mathrm{SMD}}^2}{z_{\mathrm{SCD}}^2} \iint_{A_{\mathrm{irr}}} \frac{\cos^3 \theta_A}{\pi z_{\mathrm{MCD}}^2} \, \mathrm{d}A \tag{41}$$

where z_{SMD} is the source to monitor distance, z_{SCD} is the distance from the source to the backscattering collimator surface, z_{MCD} is the monitor to backscattering surface distance, θ_A is the angle between the normal of the backscattering element dA and its view vector of the monitor and A_{irr} is the irradiated backscattering area (in the original paper the source to isocentre distance was erroneously used instead of z_{SMD} and the reflected radiation stated to be isotropic rather than diffuse). A comparison of data from the work by Lam *et al* (1998) with equation (41) yields k_b values of the order of 0.3 to 0.4 for kapton windowed chambers and approximately zero for chambers with metal sheet windows.

9.2.5. Charged particle contamination. Charged particle contamination refers to electrons released within the accelerator treatment head and in the air between the treatment head and the patient. These contribute significantly to the dose in the buildup region. The main sources of origin are (a) electrons released in traversed elements like the flattening filter, wedge etc, (b) the collimators and (c) the air. The higher the beam energy, the more dominant is the first category while air released electrons dominates at lower energies (Beauvais *et al* 1993). The lateral distribution of contaminant electrons has been modelled by a Gaussian pencil beam distribution (Nilsson and Brahme 1979, 1986, Nilsson 1985) whereas the depth dependence has been described by an exponential (Mackie 1984, Beauvais *et al* 1993, Sjögren and Karlsson 1996, Zhu and Palta 1998). Combining the Gaussian lateral dependency with the exponential depth dependency, Ahnesjö and Andreo (1989) determined the Gaussian width, the exponential attenuation and a proportionality factor from depth dose measurements by comparing with Monte Carlo generated 'clean' depth doses without electron contamination (cf section 9.1.1). The resulting parameters were used to define a pencil kernel for direct use in treatment planning (Ahnesjö *et al* 1992b).

At higher beam energies the electron contamination depth dose dependence shows a significant build up (Brahme 1987), deviating from the pure exponential commonly used. Refined measurements at lower energies have revealed a similar structure as shown by Jursinic and Mackie (1996), and they proposed the use of a biexponential function to fit all energies. Although several workers have studied the influence from different materials and their locations in the beam path (McKenna *et al* 1995, Sjögren and Karlsson 1996), no parametrized model general enough to handle all common situations has been proposed so far.

9.3. Implementation concepts

Modelling of the beam transport through the linac head is at least as complicated as modelling the transport in the patient. Hence, considerable computer time has to be spent on 'pre-patient' beam transport which is most efficiently done separately to avoid redundant calculations (several patient voxels are covered by the same beam pixel). Saxner *et al* (1997) used two energy fluence matrices, one to represent the primary beam and one to represent the head scatter fluence, thus allowing for different divergence of respective components. The head

scatter matrix was determined by integration over the scattering elements described by semi-analytical parametrizations. The scattering was treated as extended sources delineated in the calculation point of view up through the collimators (Löfgren 1998). The Peregrine project (Cox *et al* 1997) used Monte Carlo to create an output phase space file in which the position, energy and direction of the exiting particles are scored in a reference plane. The phase space file was then sampled to yield particles for further (Monte Carlo) transport in the patient. The sampling could be biased to reflect a repositioning of a collimator or similar change in the linac head geometry. For sampling efficiency, the particles were grouped into distributions depending on their origin. Liu *et al* (1997a), used a somewhat different approach for wedges and compensators. In their work, these were included within an extended phantom (as part of the patient) and their scatter contributions were predicted using kernel scaling.

10. Inverse techniques

The objective of treatment planning is to find a patient-specific optimal treatment plan that balances cure and risk for complications. It is imperative to include tools for optimization into treatment planning systems to automate the search for optimal treatments (see reviews by Brahme (1995) and Webb (1993, 1997)). Optimization of treatment plans is often referred to as 'inverse planning' since the process would start with defining treatment objectives and let the computer shape the beam set-up. Most present approaches for optimization rely on iterative searches that recomputes the dose with straight 'forward' dose calculations as reviewed above but several authors have published work that treat the dose calculation problem as an inverse problem. We will briefly review these here and also include a section on the problem to reconstruct the dose to a patient given an exit portal dose image.

10.1. Dose optimization as an inverse problem

In the early work by Brahme *et al* the shape of the lateral dose profile of a rotational x-ray beam producing a uniform, circularly symmetric dose distribution in a cylindrical phantom was expressed as a simple analytical function (Brahme *et al* 1982, Lax and Brahme 1982, Cormack and Cormack 1987, Cormack 1987). Later, a general integral equation over the irradiated volume V was expressed to illustrate the degrees of freedom in optimization (Brahme 1995):

$$D(\boldsymbol{r}) = \oiiint_V \iiint h(E, \Omega, \boldsymbol{r}, \boldsymbol{s}) f_{E,\Omega}(\boldsymbol{s})\, \mathrm{d}E\, \mathrm{d}^2\Omega\, \mathrm{d}^3 s \tag{42}$$

where $\boldsymbol{s}$ denotes the position of the kernel centre, h is a problem-specific energy deposition kernel, f a kernel density function, E the energy of the incident particle and Ω the direction of incidence. For a stationary and monoenergetic beam the above equation (24) simplifies into equation (24) where terma is the kernel density function. The idea of inverse methods for dose calculation was to design a single kernel for several beams and find direct numerical solutions to equation (42). This approach implied pre-assumptions of the beam set-up to include rotations, etc, into a problem-specific rather than radiation quality-specific kernel. One approach has been to use rotational kernels generated from rotated pencil kernels in a cylindrical phantom (Eklöf *et al* 1990, Desobry *et al* 1991). An alternative kernel definition was proposed by Holmes *et al* (1991) where the kernel was the result of photons interacting at a point-of-convergence from a number of directions, calculated as a weighted superposition of rotated monodirectional point kernels. The kernel density is typically calculated by some iterative deconvolution technique including positivity operators to avoid negative kernel density. Incident beam profiles may then be determined through backprojection analogous to image backprojection in computed tomography.

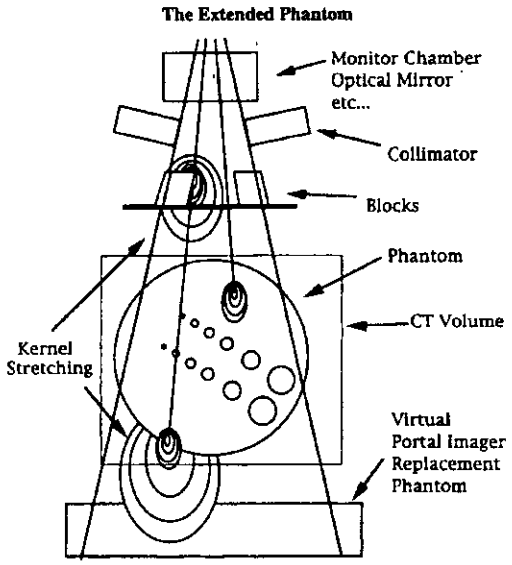


Figure 15. The 'extended phantom' used to compute the dose to a portal imager phantom. (From McNutt *et al* (1996b), with permission.)

The major limitation of the purely inverse approach is that the desired dose distribution is generally not obtainable due to the positivity requirement. Hence, more recent optimization research applies iterations using forward dose calculations (Gustafsson *et al* 1995, Mohan *et al* 1996) and focuses on objective function design (Källman *et al* 1992a, Bortfeld *et al* 1997) and also inclusion of set-up and delivery uncertainties (Löf *et al* 1998, Yan *et al* 1998).

10.2. Dose reconstruction from portal images

Portal dose images (PDIs) contain geometric and dosimetric information useful for treatment verification. Use of portal imaging is a rapidly growing area and we will restrict ourselves to applications of relevance to dose calculations. Transmission dosimetry, as it was originally called by Wong *et al* (1990), refers to the calculation of dose at the portal image plane. The concept is clearly illustrated in the work by McNutt *et al* (1996a, b) where they used an extended phantom concept (see figure 15) to let the portal imager constitute part of the irradiation geometry in a superposition dose calculation algorithm. The radiation transport physics governs the relations between the beam, the patient, the patient dose and the portal image dose thus providing a variety of possibilities to construct verification procedures based on consistency checks of calculated and measured entities.

Ying *et al* (1990) used the entrance fluence matrix and proposed an iterative method where the CT data were corrected to yield portal dose images matching measured images. McNutt *et al* (1996b) assumed, for tomotherapy applications (Mackie *et al* 1993), that the CT density matrix used in the portal dose calculation equals the one used during treatment and proposed to use the technique for dose verification. An alternative to tomotherapy to acquire a representative CT would be megavoltage CT acquisition during treatment (Swindell *et al* 1983, Brahme *et al* 1987, Lewis *et al* 1992).

3295

Hansen *et al* (1996) on the other hand, chose to perform a rigid body transform on the planning CT density map of the patient which assumes that the patient anatomy is unchanged between planning and treatment except for an overall rotation and/or translation. This transform is performed by an operator which is obtained from matching a number of electronic portal images to their corresponding digitally reconstructed radiographs. The dose reconstruction model by Hansen *et al*, unlike the work by McNutt *et al*, relies on a pencil beam kernel convolution algorithm and uses a different normalization approach to access the primary photon fluence level.

Information from an electronic portal imaging device (EPID) in conjunction with a convolution type dose calculation model has also been used to calculate patient exit surface dose distributions (Boellaard *et al* 1997). The primary dose at the imaging device, as extracted from a measured transmission dose, is projected to the exit surface of the patient. From this, the exit scatter dose is determined by convolving the primary dose with an exit dose spread function, which describes the lateral distribution of the scatter exit dose resulting from a pencil beam. The method has been applied to phantoms of variable thickness with the aid of a measured, field size-independent, renormalization correction that takes into account the variation of the scatter dose with phantom thickness. The estimation of exit dose, as the sum of primary and scatter exit doses, in this manner, works well for homogeneous media. The method, however, is probably limited in heterogeneous media mainly because the lateral extent and position of heterogeneity is not taken into account.

11. Conclusions

The need for general dose calculations has long been an issue, since treatment machines are engineered for delivery of complex treatment techniques envisaged to improve treatment results. The motivation for high dose accuracy is the steep dose response of tissues in combination with narrow therapeutic windows. The price–performance relationship for computer systems has not until recently enabled the dose calculation problem to be treated from a first-principle point. Still the most fundamental method, direct particle transport by Monte Carlo simulations is too slow to be practical in routine planning. This can be circumvented by use of parallel hardware but the additional complexity and cost is probably not justified since the point kernel models (with related beam characterization models) fulfil reasonable accuracy requirements. Monte Carlo is, however, firmly established as a necessary tool for providing basic and benchmarking data for faster methods, both for patient dose calculations and for treatment machine characterization. The consistency between experimental methods, Monte Carlo generated phase space distributions and semianalytical methods for beam characterization is an important area for further research. For routine use, point kernel models are well examined and are also conceptually simple, which is important for the understanding of dose results. Charged particle disequilibrium, a long-standing issue for lung dose calculations, is modelled by point kernel models but there is a residual error caused by the rectilinear density scaling approach. Better scaling methods have been proposed but are still premature. Pencil kernel models share the features of point kernel models with respect to beam modulations but have the same limitations as classical broad beam models with respect to the impact from heterogeneities. Pencil kernel models are, however, faster and simpler to use than point kernel models and will therefore be used for the foreseeable future as part of the optimization algorithms for finding optimal beam modulations and field shapes. Electron contamination is not addressed by the basic photon kernel models and treatment planning implementations have to rely on separate models which at present are less well developed than the pure photon models. This weakness is generally bypassed by choosing calibration procedures utilizing depth greater than the maximum depth of contamination influence.

3246

The increased complexity encountered in 3D conformal therapy implies an increased dependency on monitor unit calculations provided by the treatment planning systems. Implementation errors, modelling limitations and flaws in system handling and data logistics make independent checks of monitor unit calculations imperative. These should be simple to implement and rely on independently measured data. Suitable candidates for such models are scatter integration and scaling techniques, as reviewed in sections 5.2.1 and 5.2.2, combined with 1D depth scalings for heterogeneities. A general problem with empirical scatter scaling techniques is that they are developed for open, not modulated beams, and there is a great need to improve these models to include effects from modulations. Pencil kernel models will probably provide the link between empirical results and first principle models in the search for such methods. The need for quality assurance of complex treatment delivery has boosted an interest in electronic portal imaging. Several computation models have the potential to calculate the dose also to an imager, thus providing additional tools for dosimetric verification.

The level of accuracy achieved by any kind of model is subject to implementation details which should be clearly outlined by the manufacturers of treatment planning systems. The quality assurance of treatment related data is a tedious task requiring a cooperative attitude across organizational borders. In the end, no dose results can be better than the model approximations and the quality of data fed into the model.

Acknowledgments

The detailed and constructive criticism provided by the Board Members is gratefully acknowledged.

References

Ågren-Cronqvist A-K 1995 Quantification of the response of heterogeneous tumours and organized normal tissues to fractionated radiotherapy *PhD Thesis* University of Stockholm

Ahnesjö A 1984 Application of transform algorithms for calculation of absorbed dose in photon beams *Int. Conf. on the Use of Computers in Radiation Therapy, VIII ICCR (Toronto, Canada)* (Los Alamos, CA: IEEE Computer Society Press) pp 17–20

——1987 Invariance of convolution kernels applied to dose calculations for photon beams *Int. Conf. on the Use of Computers in Radiation Therapy, IX ICCR (Scheveningen, The Netherlands)* ed I A D Bruinvis, P H van der Giessen, H J van Kleffens and F W Wittkämper (Amsterdam: Elsevier) pp 99–102

——1989 Collapsed cone convolution of radiant energy for photon dose calculation in heterogeneous media *Med. Phys.* **16** 577–92

——1991 Dose calculation methods in photon beam therapy using energy deposition kernels *PhD Thesis* University of Stockholm

——1992 Dose calculation methods for multidimensional treatment planning *Three-Dimensional Treatment Planning (Geneva)* ed P Minet (Liège: European Association of Radiology) pp 277–88

——1994 Analytic modeling of photon scatter from flattening filters in photon therapy beams *Med. Phys.* **21** 1227–35

——1995 Collimator scatter in photon therapy beams *Med. Phys.* **22** 267–78

——1997 Cone discretization for the collapsed cone algorithm *Int. Conf. on the Use of Computers in Radiation Therapy, XII ICCR (Salt Lake City, Utah, USA)* ed D D Leavitt and G Starkschall (Madison, WI: Medical Physics Publishing) pp 114–16

Ahnesjö A and Andreo P 1989 Determination of effective bremsstrahlung spectra and electron contamination for photon dose calculations *Phys. Med. Biol.* **34** 1451–64

Ahnesjö A, Andreo P and Brahme A 1987 Calculation and application of point spread functions for treatment planning with high energy photon beams *Acta Oncol.* **26** 49–56

Ahnesjö A, Knöös T and Montelius A 1992a Application of the convolution method for calculation of output factors for therapy photon beams *Med. Phys.* **19** 295–301

Ahnesjö A and Mackie T R 1987 Analytical description of Monte Carlo generated photon dose convolution kernels *Int. Conf. on the Use of Computers in Radiation Therapy, IX ICCR (Scheveningen, The Netherlands)* ed I A D Bruinvis, P H van der Giessen, H J van Kleffens and F W Wittkämper (Amsterdam: Elsevier) pp 197–200

Ahnesjö A, Saxner M and Trepp A 1992b A pencil beam model for photon dose calculation *Med. Phys.* **19** 263–73

Ahnesjö A and Trepp A 1991 Acquisition of the effective lateral energy fluence distribution for photon beam dose calculations by convolution models *Phys. Med. Biol.* **36** 973–85

Ahnesjö A, Weber L and Nilsson P 1995 Modeling transmission and scatter for photon beam attenuators *Med. Phys.* **22** 1711–20

Alcouffe R E, Baker R S, Brinkley F W, Marr D R, O'Dell R D and Walters W F 1995 DANTSYS: a diffusion accelerated neutral particle transport code system *Report* LA-12969-M UC-705 (Los Alamos National Laboratory)

Andreo P 1990 Uncertainties in dosimetric data and beam calibration *Int. J. Radiat. Oncol. Biol. Phys.* **19** 1233–47

——1991 Monte Carlo techniques in medical radiation physics *Phys. Med. Biol.* **36** 861–920

Andreo P and Brahme A 1984 Restricted energy loss straggling and multiple scattering of electrons in mixed Monte Carlo procedures *Radiat. Res.* **100** 16–29

Archer B R, Almond P R and Wagner L K 1985 Application of a Laplace transform pair model for high-energy x-ray spectral reconstruction *Med. Phys.* **12** 630–3

Åsell M 1999 Development of optimized radiation therapy using external electron and photon beams *PhD Thesis* University of Stockholm

Aspradakis M M 1996 A study to assess and improve dose computations in photon beam therapy *PhD Thesis* University of Edinburgh

Aspradakis M M and Redpath A T 1997 A technique for the fast calculation of three-dimensional photon dose distributions using the superposition model. *Phys. Med. Biol.* **42** 1475–89

Attix F A 1986 *Introduction to Radiological Physics and Radiation Dosimetry* (New York: Wiley)

Austin-Seymour M, Chen G T Y, Rosenman J, Michalski J, Lindsley K and Goitein M 1995 Tumor and target delineation: current research and future challenges *Int. J. Radiat. Oncol. Biol. Phys.* **33** 1041–52

Baker C R, Amaée B and Spyrou N M 1995 Reconstruction of megavoltage photon spectra by attenuation analysis *Phys. Med. Biol.* **40** 529–42

Baker C R and Peck K K 1997 Reconstruction of 6 MV photon spectra from measured transmission including maximum energy estimation *Phys. Med. Biol.* **42** 2041–51

Batho H F 1964 Lung corrections in cobalt 60 beam therapy *J. Can. Assoc. Radiol.* **15** 79–83

Battista J J and Sharpe M B 1992 True three-dimensional dose computations for megavoltage x-ray therapy: a role for the superposition principle *Australas. Phys. Eng. Sci. Med.* **15** 159–78

Beaudoin L 1968 Analytical approach to the solution of the dosimetry in heterogeneous media *MSc Thesis* University of Toronto

Beauvais H, Bridier A and Dutreix A 1993 Characteristics of contamination electrons in high energy photon beams *Radiother. Oncol.* **29** 308–16

Bell G I and Glasstone S 1970 *Nuclear Reactor Theory* (Malabar, FL: Krieger)

Bellman R and Wing G M 1975 *An Introduction to Invariant Imbedding* (New York: Wiley)

Berger M J 1963 Monte Carlo calculation of the penetration and diffusion of fast charged particles *Methods in Computational Physics* vol 1, ed B Alder, S Fernbach and M Rotenberg (New York: Academic) pp 135–215

Biclajew A F and Rogers D W O 1987 PRESTA: The parameter reduced electron-step transport algorithm for electron Monte Carlo transport *Nucl. Instrum. Methods* B **18** 165–81

BJR 1983 Central axis depth dose data for use in radiotherapy *Br. J. Radiol.* (suppl 17)

——1996 Central axis depth dose data for use in radiotherapy *Br. J. Radiol.* (suppl 25)

Bjärngard B E 1987 On Fano's and O'Connor's theorems *Radiat. Res.* **109** 184–9

Bjärngard B E and Cunningham J R 1986 Comments on 'Validity of the concept of separating primary and scatter dose' (letter) *Med. Phys.* **13** 760–1

Bjärngard B E and Petti P L 1988 Description of the scatter component in photon-beam data *Phys. Med. Biol.* **33** 21–32

Bjärngard B E and Shackford H 1994 Attenuation in high-energy x-ray beams *Med. Phys.* **21** 1069–73

Bjärngard B E and Siddon R L 1982 A note on equivalent circles, squares, and rectangles *Med. Phys.* **9** 258–60

Bjärngard B E, Tsai J S and Rice R K 1989 Attenuation in very narrow photon beams *Radiat. Res.* **118** 195–200

——1990 Doses on the central axes of narrow 6-MV x-ray beams *Med. Phys.* **17** 794–9

Bjärngard B E and Vadash P 1995 Analysis of central-axis doses for high-energy x-rays *Med. Phys.* **22** 1191–5

Bjärngard B E, Vadash P and Ceberg C P 1997 Quality control of measured x-ray beam data *Med. Phys.* **24** 1441–4

Bloch P and Altschuler M D 1995 Three-dimensional photon beam calculations *Radiation Therapy Physics* ed A Smith (Berlin: Springer)

Boellaard B E, van Herk M and Mijnheer B J 1997 A convolution model to convert transmission dose images to exit dose distributions *Med. Phys.* **24** 189–99

Börgers C 1998 Complexity of Monte Carlo and deterministic dose-calculation methods *Phys. Med. Biol.* **43** 517–28

Bortfeld T, Boyer A L, Schlegel W, Kahler D L and Waldron T J 1994 Realization and verification of three-dimensional conformal radiotherapy with modulated fields *Int. J. Radiat. Oncol. Biol. Phys.* **30** 899–908

Bortfeld T, Schlegel W and Rhein B 1993 Decomposition of pencil beam kernels for fast dose calculations in three-dimensional treatment planning *Med. Phys.* **20** 311–18

Bortfeld T, Stein J and Preiser K 1997 Clinically relevant intensity modulation optimization using physical criteria *Int. Conf. on the Use of Computers in Radiation Therapy, XII ICCR (Salt Lake City, Utah, USA)* ed D D Leavitt and G Starkschall (Madison, WI: Medical Physics Publishing) pp 1–4

Bourland J D and Chaney E L 1992 A finite-size pencil beam model for photon dose calculations in three dimensions *Med. Phys.* **19** 1401–12

Boyer A L 1988 Relationship between attenuation coefficients and dose-spread kernels *Radiat. Res.* **113** 235–42

Boyer A L and Mok E C 1984 Photon beam modeling using Fourier transform techniques *Int. Conf. on the Use of Computers in Radiation Therapy, VIII ICCR (Toronto, Canada)* (Los Alamos, CA: IEEE Computer Society Press) pp 14–16

——1985 A photon dose distribution model employing convolution calculations *Med. Phys.* **12** 169–77

——1986 Calculation of photon dose distributions in an inhomogeneous medium using convolutions *Med. Phys.* **13** 503–9

Boyer A L and Schultheiss T 1988 Effects of dosimetric and clinical uncertainty on complication-free local tumor control *Radiother. Oncol.* **11** 65–71

Boyer A L, Zhu Y, Wang L and Francois P 1989 Fast Fourier transform convolution calculations of x-ray isodose distributions in homogeneous media *Med. Phys.* **16** 248–53

Brahme A 1977 Restricted mass energy absorption coefficients for use in dosimetry *Internal Report* SSI:1977-009 (National Institute of Radiation Protection)

——1984 Dosimetric precision requirements in radiation therapy *Acta Radiol. Oncol.* **23** 379–91

——1987 Design principles and clinical possibilities with a new generation of radiation therapy equipment. A review *Acta Oncol.* **26** 403–12

——(ed) 1988 Accuracy requirements and quality assurance of external beam therapy with photons and electrons *Acta Radiol. Oncol.* (suppl 1)

——1995 Treatment optimization using physical and radiobiological objective functions *Radiation Therapy Physics* ed A Smith (Berlin: Springer)

Brahme A, Lind B and Nafstadius P 1987 Radiotherapeutic computed tomography with scanned photon beams *Int. J. Radiat. Oncol. Biol. Phys.* **13** 95–101

Brahme A, Roos J E and Lax I 1982 Solution of an integral equation encountered in rotation therapy *Phys. Med. Biol.* **27** 1221–9

Briesmeister J F 1988 MCNP—A general Monte Carlo *N*-particle transport code *Report* LA-12625-M (Los Alamos National Laboratory)

Cassell K J, Hobday P A and Parker R P 1981 The implementation of a generalised Batho inhomogeneity correction for radiotherapy planning with direct use of CT numbers *Phys. Med. Biol.* **26** 825–33

Ceberg C P, Bjärngard B E and Zhu T C 1996 Experimental determination of the dose kernel in high-energy x-ray beams *Med. Phys.* **23** 505–11

Chaney E L, Cullip T J and Gabriel T A 1994 A Monte Carlo study of accelerator head scatter *Med. Phys.* **21** 1383–90

Chin L M, Kijewski P K, Svensson G K and Bjärngard B E 1983 Dose optimization with computer-controlled gantry rotation, collimator motion and dose-rate variation *Int. J. Radiat. Oncol. Biol. Phys.* **9** 723–9

Chui C and Mohan R 1984 Differential pencil beam dose computation model (abstract) *Med. Phys.* **11** 392

Chui C S, LoSasso T and Spirou S 1994 Dose calculation for photon beams with intensity modulation generated by dynamic jaw or multileaf collimations *Med. Phys.* **21** 1237–44

Clarkson J R 1941 A note on depth doses in fields of irregular shape *Br. J. Radiol.* **14** 265–8

Convery D J and Rosenbloom M E 1992 The generation of intensity-modulated fields for conformal radiotherapy by dynamic collimation *Phys. Med. Biol.* **37** 1359–74

Cormack A M 1987 A problem in rotation therapy with X-rays *Int. J. Radiat. Oncol. Biol. Phys.* **13** 623–30

Cormack A M and Cormack R A 1987 A problem in rotation therapy with X-rays: dose distributions with an axis of symmetry *Int. J. Radiat. Oncol. Biol. Phys.* **13** 1921–5

Cox L J, Schach von Wittenau A E, Bergstrom P M J, Mohan R, Libby B, Wu Q and Lovelock D M J 1997 Photon beam description in PEREGRINE for Monte Carlo dose calculations *Int. Conf. on the Use of Computers in Radiation Therapy, XII ICCR (Salt Lake City, Utah, USA)* ed D D Leavitt and G Starkschall (Madison, WI: Medical Physics Publishing) pp 142–5

Cunningham J R 1972 Scatter-air ratios *Phys. Med. Biol.* **17** 42–51

Cunningham J R 1983 Current and future developement of tissue inhomogeneity corrections for photon beam clinical dosimetry with the use of CT *Computed Tomography in Radiation Therapy* ed C C Ling, C C Rogers and R J Morton (New York: Raven Press) pp 209–18

Cunningham J R and Beaudoin L 1973 Calculations for tissue inhomogeneities with experimental verification *XIII Int. Congress of Radiology (Madrid)* XIII, pp 653–7

Cunningham J R, Shrivastava P N and Wilkinson J M 1972 Program IRREG-calculation of dose from irregularly shaped radiation beams *Comput. Programs Biomed.* **2** 192–9

Dahlin H, Lamm I-L, Landberg T, Levernes S and Ulsø N 1983 User requirements on CT-based computed dose planning systems in radiation therapy *Acta Radiol. Oncol.* **22** 398–414

Dasu A and Denekamp J 1999 Superfractionation as a potential hypoxic cell radiosensitizer: prediction of an optimum dose per fraction *Int. J. Radiat. Oncol. Biol. Phys.* **43** 1083–94

Day M J 1950 A note on the calculation of dose in x-ray fields *Br. J. Radiol.* **23** 368–9

——1972 The equivalent field method for axial dose determinations in rectangular fields *Br. J. Radiol.* (suppl 10)

——1978 The equivalent field method for axial dose determinations in rectangular fields *Br. J. Radiol.* (suppl 11) 95–100

——1983 The normalised peak scatter factor and normalised scatter functions for high energy photon beams *Br. J. Radiol.* (suppl 17) 131–6

Day M J and Aird E G A 1996 The equivalent field method for dose determination in rectangular fields *Br. J. Radiol.* (suppl 25) 138–51

Dean R D 1980 A scattering kernel for use in true three-dimensional dose calculations (abstract) *Med. Phys.* **7** 429

DeMarco J J, Solberg T D and Smathers J B 1998 A CT-based Monte Carlo simulation tool for dosimetry planning and analysis *Med. Phys.* **25** 1–11

Desobry G E and Boyer A L 1991 Bremsstrahlung review: an analysis of the Schiff spectrum *Med. Phys.* **18** 497–505

——1994 An analytic calculation of the energy fluence spectrum of a linear accelerator *Med. Phys.* **21** 1943–52

Desobry G E, Wells N H and Boyer A L 1991 Rotational kernels for conformal therapy *Med. Phys.* **18** 481–7

Difilippo F C 1998 Forward and adjoint methods for radiotherapy planning *Med. Phys.* **25** 1702–10

Dische S, Saunders M I, Williams C, Hopkins A and Aird E 1993 Precision in reporting the dose given in a course of radiotherapy *Radiother. Oncol.* **29** 287–93

Dunscombe P B and Nieminen J M 1992 On the field-size dependence of relative output from a linear accelerator *Med. Phys.* **19** 1441–4

Dutreix A, Bjärngard B E, Bridier A, Mijnheer B, Shaw J E and Svensson H 1997 Monitor unit calculation for high energy photon beams *Physics for Clinical Radiotherapy, Booklet No. 3* (Leuven/Apeldoorn: Garant)

Dutreix J, Dutreix A and Tubiana M 1965 Electronic equilibrium and transition stages *Phys. Med. Biol.* **10** 177–90

Ebert M A, Hoban P W and Keall P J 1996 Modelling clinical accelerator beams: a review *Australas. Phys. Eng. Sci. Med.* **19** 131–50

Eklöf A, Ahnesjö A and Brahme A 1990 Photon beam energy deposition kernels for inverse radiotherapy planning *Acta Oncol.* **29** 447–54

El-Khatib E and Battista J J 1984 Improved lung dose calculation using tissue-maximum ratios in the Batho correction *Med. Phys.* **11** 279–86

Faddegon B A, Ross C K and Rogers D W O 1990 Forward-directed bremsstrahlung of 10- to 30-MeV electrons incident on thick targets of Al and Pb *Med. Phys.* **17** 773–85

——1991 Angular distribution of bremsstrahlung from 15-MeV electrons incident on thick targets of Be, Al, and Pb *Med. Phys.* **18** 727–39

Fano U 1954 Note on the Bragg-Gray cavity principle for measuring energy dissipation *Radiat. Res.* **1** 237–40

Fraass B, Doppke K, Hunt M, Kutcher G, Starkschall G, Stern R and Van Dyke J 1998 American Association of Physicists in Medicine Radiation Therapy Committee Task Group 53: quality assurance for clinical radiotherapy treatment planning *Med. Phys.* **25** 1773–829

Fraass B A 1995 The development of conformal radiation therapy *Med. Phys.* **22** 1911–21

Francois P, Coste F, Bonnet J and Caselles O 1997 Validation of reconstructed bremsstrahlung spectra between 6 MV and 25 MV from measured transmission data *Med. Phys.* **24** 769–73

Goitein M and Abrams M 1983 Multi-dimensional treatment planning: I. Delineation of anatomy *Int. J. Radiat. Oncol. Biol. Phys.* **9** 777–87

Goitein M, Abrams M, Rowell D, Pollari H and Wiles J 1983 Multi-dimensional treatment planning: II. Beam's eye-view, back projection, and projection through CT sections *Int. J. Radiat. Oncol. Biol. Phys.* **9** 789–97

Gokhale P, Hussein E M A and Kulkarni N 1994 Determination of beam orientation in radiotherapy planning *Med. Phys.* **21** 393–400

Greene D and Stewart J R 1965 Isodose curves in non-uniform phantoms *Br. J. Radiol.* **38** 378–85

Gupta S K and Cunningham J R 1966 Measurement of tissue-air ratios and scatter functions for large field sizes, for cobalt 60 gamma radiation *Br. J. Radiol.* **39** 7–11

Gustafsson A 1996 Development of a versatile algorithm for optimization of radiation therapy *PhD Thesis* University of Stockholm

Gustafsson A, Lind B K and Brahme A 1994 A generalized pencil beam algorithm for optimization of radiation therapy *Med. Phys.* **21** 343–56

Gustafsson A, Lind B K, Svensson R and Brahme A 1995 Simultaneous optimization of dynamic multileaf collimation and scanning patterns or compensation filters using a generalized pencil beam algorithm *Med. Phys.* **22** 1141–56

Hannallah D, Zhu T C and Bjärngard B E 1996 Electron disequilibrium in high-energy x-ray beams *Med. Phys.* **23** 1867–71

Hansen V N, Evans P M and Swindell W 1996 The application of transit dosimetry to precision radiotherapy *Med. Phys.* **23** 713–21

Harder D 1974 Fano's theorem and the multiple scattering correction. *Fourth Symp. on Microdosimetry (Verbania-Pallanza, Italy)* ed J Booz, H G Ebert, R Eickel and A Waker (Luxembourg: Commission of the European Communities) pp 677–93

Harms W B, Low D A, Wong J W and Purdy J A 1998 A software tool for the quantitative evaluation of 3D dose calculation algorithms *Med. Phys.* **25** 1830–36

Hartmann Siantar C L, Bergstrom P M, Chandler W P, Chase L, Cox L J, Daly T P, Garrett D, Hornstein S M, House R K, Moses E I, Patterson R W, Rathkopf J A and Schach von Wittenau A E 1997 Lawrence Livermore National Laboratory's PEREGRINE project *Int. Conf. on the Use of Computers in Radiation Therapy, XII ICCR (Salt Lake City, Utah, USA)* ed D D Leavitt and G Starkschall (Madison, WI: Medical Physics Publishing) pp 19–22

Heukelom S, Lanson J H and Mijnheer B J 1994 Wedge factor constituents of high energy photon beams: field size and depth dependence *Radiother. Oncol.* **30** 66–73

——1997 Differences in wedge factor determination in air using a PMMA mini-phantom or a brass buildup cap *Med. Phys.* **24** 1986–91

Hoban P W 1995 Accounting for the variation in collision kerma-to-terma ratio in polyenergetic photon beam convolution *Med. Phys.* **22** 2035–44

Hoban P W, Murray D C and Round W H 1994 Photon beam convolution using polyenergetic energy deposition kernels *Phys. Med. Biol.* **39** 669–85

Holmes T, Mackie T R, Simpkin D and Reckwerdt P 1991 A unified approach to the optimization of brachytherapy and external beam dosimetry *Int. J. Radiat. Oncil. Biol. Phys.* **20** 859–73

Holt J G, Laughlin J S and Moroney J P 1970 The extension of the concept of tissue-air ratios (TAR) to high-energy x-ray beams *Radiology* **96** 437–46

Hounsell A R 1998 Monitor chamber backscatter for intensity modulated radiation therapy using multileaf collimators *Phys. Med. Biol.* **43** 445–54

Huang P H, Kase K R and Bjärngard B E 1983 Reconstruction of 4-MV bremsstrahlung spectra from measured transmission data *Med. Phys.* **10** 778–85

Huizenga H and Storchi P R M 1989 Numerical calculation of energy deposition by broad high-energy electron beams *Phys. Med. Biol.* **34** 1371–96

Hurkmans C, Knöös T, Nilsson P, Svahn-Tapper G and Danielsson H 1995 Limitations of a pencil beam approach to photon dose calculations in the head and neck region *Radiother. Oncol.* **37** 74–80

IAEA 1987 Absorbed dose determination in photon and electron beams. An international code of practice *IAEA Technical Report Series* No. 277 (Vienna: International Atomic Energy Agency)

ICCR 1970 Computers in radiotherapy *Int. Conf. on the Use of Computers in Radiation Therapy, III ICCR (Glasgow, Scotland)* ed A S Glicksman, M Cohen and J R Cunningham (British Journal of Radiology)

ICRU 1963 Clinical dosimetry handbook 87 *ICRU Publication* 10d (Washington, DC: National Bureau of Standards)

——1973 Measurement of absorbed dose in a phantom irradiated by a single beam of x or gamma rays *ICRU Publication* 23 (Bethesda, MA: International Commission on Radiation Units and Measurements)

——1976 Determination of absorbed dose in a patient irradiated by beams of x or gamma rays in radiotherapy procedures *ICRU Publication* 24 (Bethesda, MD: International Commission on Radiation Units and Measurements)

——1987 Use of computers in external beam radiotherapy procedures with high-energy photons and electrons *ICRU Publication* 42 (Bethesda, MD: International Commission on Radiation Units and Measurements)

——1998 Fundamental quantities and units for ionizing radiation *ICRU Publication* 60 (Bethesda, MD: International Commission on Radiation Units and Measurements)

Islam M K and Van Dyk J 1995 Effects of scatter generated by beam-modifying absorbers in megavoltage photon beams *Med. Phys.* **22** 2075–81

Iwasaki A 1985 A method of calculating high-energy photon primary absorbed dose in water using forward and backward spread dose-distribution functions *Med. Phys.* **12** 731–7

——1990 Calculation of three-dimensional photon primary absorbed dose using forward and backward spread dose-distribution functions *Med. Phys.* **17** 195–202

Iwasaki A and Ishito T 1984 The differential scatter-air ratio and differential backscatter factor method combined with the density scaling theorem *Med. Phys.* **11** 755–63

Jaffray D A, Battista J J, Fenster A and Munro P 1993 X-ray sources of medical linear accelerators: focal and extra-focal radiation *Med. Phys.* **20** 1417–27

Janssen J J, Korevaar E W, Storchi P R M and Huizenga H 1997 Numerical calculation of energy deposition by high-energy electron beams: III-B. Improvements to the 6D phase space evolution model *Phys. Med. Biol.* **42** 1441–9

Johns H E, Darby E K, Haslam R N, Katz L and Harrington E L 1949 Depth dose data and isodose distributions for radiation from a 22 MeV betatron *Am. J. Roentgenol.* **62** 257–68

Johns H E, Whitmore G F, Watson T A and Umberg F H 1953 A system of dosimetry for rotation therapy and typical rotation distribution *J. Can. Assoc. Radiol.* **4** 1–14

Jung B, Montelius A, Dahlin H, Ekström P, Ahnesjö A, Högström B and Glimelius B 1997 The conceptual design of a radiation oncology planning system *Comput. Methods Prog. Biom.* **52** 79–92

Jursinic P A and Mackie T R 1996 Characteristics of secondary electrons produced by 6, 10 and 24 MV x-ray beams *Phys. Med. Biol.* **41** 1499–509

Karlsson M, Nyström H and Svensson H 1993 Photon beam characteristics on the MM50 racetrack microtron and a new approach for beam quality determination *Med. Phys.* **20** 143–9

Karzmark C J, Deubert A and Loevinger R 1965 Letter to Editor: tissue-phantom ratios—an aid to treatment planning *Br. J. Radiol.* **38** 158–9

Kase K R and Svensson G K 1986 Head scatter data for several linear accelerators (4–18 MV) *Med. Phys.* **13** 530–2

Kawrakow I and Bielajew A F 1998 On the condensed history technique for electron transport *Nucl. Instrum. Methods* B **142** 253–80

Kawrakow I, Fippel M and Friedrich K 1996 3D electron dose calculation using a voxel based Monte Carlo algorithm (VMC) *Med. Phys.* **23** 445–57

Keall P and Hoban P 1995 Accounting for primary electron scatter in x-ray beam convolution calculations *Med. Phys.* **22** 1413–18

——1996a Superposition dose calculation incorporating Monte Carlo generated electron track kernels *Med. Phys.* **23** 479–85

——1996b A review of electron beam dose calculation algorithms *Australas. Phys. Eng. Sci. Med.* **19** 111–30

Keller H, Fix M and Rüegsegger P 1998 Calibration of a portal imaging device for high-precision dosimetry: a Monte Carlo study. *Med. Phys.* **25** 1891–902

Khan F M, Moore V C and Sato S 1972 Depth dose and scatter analysis of 10 MV X rays *Radiology* **102** 165–9

Kijewski P K, Bjärngard B E and Petti P L 1986 Monte Carlo calculations of scatter dose for small field sizes in a ^{60}Co beam *Med. Phys.* **13** 74–7

Kijewski P K, Chin L M and Bjärngard B E 1978 Wedge-shaped dose distributions by computer-controlled collimator motion *Med. Phys.* **5** 426–9

Kim S, Palta J R and Zhu T C 1998 A generalized solution for the calculation of in-air output factors in irregular fields *Med. Phys.* **25** 1692–701

King L V 1912 Absorption problems in radioactivity *Phil. Mag.* **xxiii** 242–50

Knöös T, Ahnesjö A, Nilsson P and Weber L 1995 Limitations of a pencil beam approach to photon dose calculations in lung tissue *Phys. Med. Biol.* **40** 1411–20

Koch H W and Motz J W 1959 Bremsstrahlung cross-section formulas and related data *Rev. Mod. Phys.* **31** 920–55

Kubo H 1989 Telescopic measurements of backscattered radiation from secondary collimator jaws to a beam monitor chamber using a pair of slits *Med. Phys.* **16** 295–8

Küster G, Bortfeld T and Schlegel W 1997 Monte Carlo simulations of radiation beams from radiotherapy units and beam limiting devices using the program GEANT *Int. Conf. on the Use of Computers in Radiation Therapy, XII ICCR (Salt Lake City, Utah, USA)* ed G Starkschall (Madison, WI: Medical Physics Publishing) pp 150–2

Källman P, Ågren A and Brahme A 1992b Tumour and normal tissue responses to fractionated non-uniform dose delivery. *Int. J. Radiat. Biol.* **62** 249–62

Källman P, Lind B K and Brahme A 1992a An algorithm for maximizing the probability of complication-free tumour control in radiation therapy *Phys. Med. Biol.* **37** 871–90

Lam K L, Muthuswamy M S and Ten Haken R K 1998 Measurement of backscatter to the monitor chamber of medical accelerators using target charge *Med. Phys.* **25** 334–8

Larson K B and Prasad S C 1978 Absorbed dose computations for inhomogeneous media in radiation-treatment planning using differential scatter-air ratios *2nd Ann. Symp. on Computer Applications in Medical Care (Washington DC)* (New York: IEEE) pp 93–9

Lax I and Brahme A 1982 Rotation therapy using a novel high-gradient filter *Radiology* **145** 473–8

Lee P C 1997 Monte Carlo simulations of the differential beam hardening effect of a flattening filter on a therapeutic x-ray beam *Med. Phys.* **24** 1485–9

Lewis D G, Swindell W, Morton E J, Evans P M and Xiao Z R 1992 A megavoltage CT scanner for radiotherapy verification *Phys. Med. Biol.* **37** 1985–99

Liu H H, Mackie T R and McCullough E C 1997a Calculating dose and output factors for wedged photon radiotherapy fields using a convolution/superposition method *Med. Phys.* **24** 1714–28

——1997b Calculating output factors for photon beam radiotherapy using a convolution/superposition method based on a dual source photon beam model *Med. Phys.* **24** 1975–85

——1997c Correcting kernel tilting and hardening in convolution/superposition dose calculations for clinical divergent and polychromatic photon beams *Med. Phys.* **24** 1729–41

——1997d A dual source photon beam model used in convolution/superposition dose calculations for clinical megavoltage x-ray beams *Med. Phys.* **24** 1960–74

Liu H H, McCullough E C and Mackie T R 1998 Calculating dose distributions and wedge factors for photon treatment fields with dynamic wedges based on a convolution/superposition method *Med. Phys.* **25** 56–63

Loevinger R 1950 Distribution of absorbed energy around a point source of beta radiation *Science* **112** 530–1

——1956 The dosimetry of beta sources in tissue: the point source function *Radiology* **66** 55–62

Löf J, Lind B K and Brahme A 1998 An adaptive control algorithm for optimization of intensity modulated radiotherapy considering uncertainties in beam profiles, patient set-up and internal organ motion. *Phys. Med. Biol.* **43** 1605–28

Löfgren A 1998 Polygon delineation of thick collimator apertures from arbitrary view points *MSc Thesis* University of Uppsala

Lovelock D M, Chui C S and Mohan R 1995 A Monte Carlo model of photon beams used in radiation therapy *Med. Phys.* **22** 1387–94

Luxton G and Astrahan M A 1988 Output factor constituents of a high-energy photon beam *Med. Phys.* **15** 88–91

Lydon J M 1998 Photon dose calculations in homogeneous media for a treatment planning system using a collapsed cone superposition convolution algorithm *Phys. Med. Biol.* **43** 1813–22

Mackie T R 1984 A study of charged particles and scattered photons in mega-voltage x-ray beams *PhD Thesis* University of Alberta

——1990 Applications of the Monte Carlo method in radiotherapy *The Dosimetry of Ionizing Radiation* vol 3, ed K R Kase, B Bjärngard and F H Attix (New York: Academic) pp 541-620

Mackie T R, Bielajew A F, Rogers D W O and Battista J J 1988 Generation of photon energy deposition kernels using the EGS Monte Carlo code *Phys. Med. Biol.* **33** 1–20

Mackie T R, Holmes T, Swerdloff S, Reckwerdt P, Deasy J O, Yang J, Paliwal B and Kinsella T 1993 Tomotherapy: a new concept for the delivery of dynamic conformal radiotherapy *Med. Phys.* **20** 1709–19

Mackie T R, Reckwerdt P, McNutt T, Gehring M and Sanders C 1996 Photon beam dose computations *Teletherapy: Present and Future* ed J Palta and T R Mackie (College Park, MD: American Association of Physicists in Medicine) pp 103–35

Mackie T R, Reckwerdt P and Papanikolaou N 1995 3-D photon beam dose algorithms *3-D Radiation Treatment Planning and Conformal Therapy* ed J A Purdy and B Emami (Madison, WI: Medical Physics Publishing)

Mackie T R and Scrimger J W 1984 Computing radiation dose for high energy X-rays using a convolution method *Int. Conf. on the Use of Computers in Radiation Therapy, VIII ICCR (Toronto, Canada)* (Los Alamos, CA: IEEE Computer Society Press) pp 36–40

Mackie T R, Scrimger J W and Battista J J 1985 A convolution method of calculating dose for 15-MV x rays *Med. Phys.* **12** 188–96

Manfredotti C, Nastasi U, Marchisio R, Ongaro C, Gervino G, Ragona R, Anglesio S and Sannazzari G 1990 Monte Carlo simulation of dose distribution in electron beam radiotherapy treatment planning *Nucl. Instrum. Methods* A **291** 646–54

Mayneord W V 1945 Energy absorption. IV The mathematical theory of integral dose in radium therapy *Br. J. Radiol.* **18** 12–19

McGary J E and Boyer A L 1997 An interactive, parallel, three-dimensional fast Fourier transform convolution dose calculation using a supercomputer *Med. Phys.* **24** 519–22

McKenna M G, Chen X G, Altschuler M D and Bloch P 1995 Calculation of the dose in the buildup region for high energy photon beam. Treatment planning when beam spoilers are employed *Radiother. Oncol.* **34** 63–8

McLellan J, Papiez L, Sandison G A, Huda W and Therrien P 1992 A numerical method for electron transport calculations *Phys. Med. Biol.* **37** 1109–25

McNutt T R, Mackie T R, Reckwerdt P and Paliwal B R 1996a Modeling dose distributions from portal dose images using the convolution/superposition method *Med. Phys.* **23** 1381–92

McNutt T R, Mackie T R, Reckwerdt P, Papanikolaou N and Paliwal B R 1996b Calculation of portal dose using the convolution/superposition method *Med. Phys.* **23** 527–35

Metcalfe P E, Hoban P W, Murray D C and Round W H 1989 Modelling polychromatic high energy photon beams by superposition *Australas. Phys. Eng. Sci. Med.* **12** 138–48

Metcalfe P E, Hoban P W, Murray D C and Round W H 1990 Beam hardening of 10 MV radiotherapy x-rays: analysis using a convolution/superposition method *Phys. Med. Biol.* **35** 1533–49

Mijnheer B J, Battermann J J and Wambersie A 1987 What degree of accuracy is required and can be achieved in photon and neutron therapy? *Radiother. Oncol.* **8** 237–52

Mohan R and Chui C S 1985 Validity of the concept of separating primary and scatter dose *Med. Phys.* **12** 726–30

——1987 Use of fast Fourier transforms in calculating dose distributions for irregularly shaped fields for three-dimensional treatment planning *Med. Phys.* **14** 70–7

Mohan R, Chui C and Lidofsky L 1985 Energy and angular distributions of photons from medical linear accelerators *Med. Phys.* **12** 592–7

——1986 Differential pencil beam dose computation model for photons *Med. Phys.* **13** 64–73

Mohan R, Chui C, Miller D and Laughlin J S 1981 Use of computerized tomography in dose calculations for radiation treatment planning *J. Comput. Tomogr.* **5** 273–82

Mohan R, Wu Q, Wang X and Stein J 1996 Intensity modulation optimization, lateral transport of radiation, and margins *Med. Phys.* **23** 2011–21

Murray D C, Hoban P W, Metcalfe P E and Round W H 1989 3-D superposition for radiotherapy treatment planning using fast Fourier transforms *Australas. Phys. Eng. Sci. Med.* **12** 128–37

Murray D C, Hoban P W, Round W H, Graham I D and Metcalfe P E 1991 Superposition on a multicomputer system *Med. Phys.* **18** 468–73

Nelson W R, Hirayama H and Rogers D W O 1985 The EGS4 code system *Stanford Linear Accelerator Report* 265

NEMA 1998 *Digital Imaging and Communication in Medicine (DICOM)* PS 3.1-3.14 (National Electrical Manufacturers Association)

Neuenschwander H, Mackie T R and Reckwerdt P J 1995 MMC—a high-performance Monte Carlo code for electron beam treatment planning *Phys. Med. Biol.* **40** 543–74

Niemierko A and Goitein M 1990 Random sampling for evaluating treatment plans *Med. Phys.* **17** 753–62

Nigg D W, Randolph P D and Wheeler F J 1991 Demonstration of three-dimensional deterministic radiation transport theory dose distribution analysis for boron neutron capture therapy *Med. Phys.* **18** 43–53

Nilsson B 1985 Electron contamination from different materials in high energy photon beams *Phys. Med. Biol.* **30** 139–51

Nilsson B and Brahme A 1979 Absorbed dose from secondary electrons in high energy photon beams *Phys. Med. Biol.* **24** 901–12

——1981 Contamination of high-energy photon beams by scattered photons *Strahlentherapie* **157** 181–6

——1986 Electron contamination from photon beam collimators *Radiother. Oncol.* **5** 235–44

Nilsson M and Knöös T 1992 Application of the Fano theorem in inhomogeneous media using a convolution algorithm *Phys. Med. Biol.* **37** 69–83

Nisbet A, Weatherburn H, Fenwick J D and McVey G 1998 Spectral reconstruction of clinical megavoltage photon beams and the implications of spectral determination on the dosimetry of such beams *Phys. Med. Biol.* **43** 1507–21

Nizin P S 1993 Electronic equilibrium and primary dose in collimated photon beams *Med. Phys.* **20** 1721–9

Nordell B and Brahme A 1984 Angular distribution and yield from bremsstrahlung targets *Phys. Med. Biol.* **29** 797–810

O'Connor J E 1957 The variation of scattered x-rays with density in an irradiated body *Phys. Med. Biol.* **1** 352–69

O'Connor J E and Malone D E 1989 A cobalt-60 primary dose spread array derived from measurements *Phys. Med. Biol.* **34** 1029–42

Ostapiak O Z, Zhu Y and Van Dyk J 1997 Refinements of the finite-size pencil beam model of three-dimensional photon dose calculation *Med. Phys.* **24** 743–50

Overgaard J and Bartelink H 1995 About tolerance and quality. An important notice to all radiation oncologists *Radiother. Oncol.* **35** 1–3

Papanikolaou N, Mackie T R, Meger-Wells C, Gehring M and Reckwerdt P 1993 Investigation of the convolution method for polyenergetic spectra *Med. Phys.* **20** 1327–36

Patomaki L K 1968 The equivalent field principle and its use in beam therapy dose calculations *Br. J. Radiol.* **41** 381–3

Petti P L, Rice R K, Mijnheer B J, Chin L M and Bjärngard B E 1987 A heterogeneity model for photon beams incorporating electron transport *Med. Phys.* **14** 349–54

Piermattei A, Arcovito G, Azario L, Bacci C, Bianciardi L, De Sapio E and Giacco C 1990 A study of quality of bremsstrahlung spectra reconstructed from transmission measurements *Med. Phys.* **17** 227–33

Purdy J A 1992 Photon dose calculations for three-dimensional radiation treatment planning *Semin. Radiat. Oncol.* **2** 235–45

Purdy J A and Prasad S C 1983 Current methods and algorithms in radiation absorbed dose calculation and the role of computed tomography: a review *Computed Tomography in Radiation Therapy* ed C C Ling, C C Rogers and R J Morton (New York: Raven Press) pp 187–97

Purdy J A, Wong J W, Harms W B, Emami B and Matthews J W 1987 State of the art of high energy photon treatment planning *Front. Radiat. Therapy Oncol.* **21** 4–24

Rathee S, McClean B A and Field C 1993 An improved method for rebinning kernels from cylindrical to Cartesian coordinates. *Med. Phys.* **20** 1343–51

Reckwerdt P J and Mackie T R 1992 Superposition/convolution speed improvements using run-length raytracing (AAPM abstract) *Med. Phys.* **19** 784

Redpath A T 1995 A beam model for 3-dimensional radiotherapy planning *Br. J. Radiol.* **68** 1356–63

Redpath A T and Thwaites D I 1991 A 3-dimensional scatter correction algorithm for photon beams *Phys. Med. Biol.* **36** 779–98

Roesch W M C 1958 Dose for nonelectronic equilibrium conditions *Radiat. Res.* **9** 399–410

Rogers D W O and Bielajew A F 1990 Monte Carlo techniques of electron and photon transport for radiation *The Dosimetry of Ionizing Radiation* vol 3, ed K R Kase, B Bjärngard and F H Attix (New York: Academic) pp 427–539

Rogers D W O, Faddegon B A, Ding G X, Ma C M, We J and Mackie T R 1995 BEAM: a Monte Carlo code to simulate radiotherapy treatment units *Med. Phys.* **22** 503–24

Rossi H H and Roesch W C 1962 Field equations in dosimetry *Radiat. Res.* **16** 783–95

Sätherberg A, Karlsson M and Karlsson M 1996 Theoretical and experimental determination of phantom scatter factors for photon fields with different radial energy variation *Phys. Med. Biol.* **41** 2687–94

Sauer O A 1994 Dosisverteilungen an Material-Grenzflächen bei energiereichen Röntgenstrahlen *PhD Thesis* University of Julius-Maximilians Würzburg

——1995 Calculation of dose distributions in the vicinity of high-Z interfaces for photon beams *Med. Phys.* **22** 1685–90

Sauer O and Neumann M 1990 Reconstruction of high-energy bremsstrahlung spectra by numerical analysis of depth-dose data *Radiother. Oncol.* **18** 39–47

Saxner M and Ahnesjö A 1998 Implementation of the collapsed cone method for clinical beam qualities *Med. Phys.* **25** A185

Saxner M, Löfgren A and Ahnesjö A 1997 Integration of head scatter fluence calculations in treatment planning *Int. Conf. on the Use of Computers in Radiation Therapy, XII ICCR (Salt Lake City, Utah, USA)* ed D D Leavitt and G Starkschall (Madison, WI: Medical Physics Publishing) pp 213–15

Schad L R, Gademann G, Knopp M, Zabel H J, Schlegel W and Lorenz W J 1992 Radiotherapy treatment planning of basal meningiomas: improved tumor localization by correlation of CT and MR imaging data *Radiother. Oncol.* **25** 56–62

Schoknecht G 1971 Die Beschreibung von Strahlenfeldern durch Separierung von Primär- und Streustrahlung IV. Berechnung von Streuverteilungen für parallele Photonenstrahlenfelder *Strahlentherapie* **141** 326–31

Shapiro A, Lin B I, Windham J P and Kereiakes J G 1976 Transport calculations of gamma ray flux density and dose rate about implantable californium-252 sources *Phys. Med. Biol.* **21** 509–23

Sharpe M B and Battista J J 1993 Dose calculations using convolution and superposition principles: the orientation of dose spread kernels in divergent x-ray beams *Med. Phys.* **20** 1685–94

Sharpe M B, Jaffray D A, Battista J J and Munro P 1995 Extrafocal radiation: a unified approach to the prediction of beam penumbra and output factors for megavoltage x-ray beams *Med. Phys.* **22** 2065–74

Siddon R L 1985 Fast calculation of the exact radiological path for a three-dimensional CT array *Med. Phys.* **12** 252–5

Siddon R L, Dewyngaert J K and Bjärngard B E 1985 Scatter integration with right triangular fields *Med. Phys.* **12** 229–31

Silver M D 1994 Target self-attenuation extension to the Desobry and Boyer thick-target bremsstrahlung spectrum *Med. Phys.* **21** 577–9

Sjögren R and Karlsson M 1996 Electron contamination in clinical high energy photon beams *Med. Phys.* **23** 1873–81

Söderström S, Gustafsson A and Brahme A 1993 The clinical value of different treatment objectives and degrees of freedom in radiation therapy optimization *Radiother. Oncol.* **29** 148–63

Sontag M R 1979 Photon beam dose calculations in regions of tissue heterogeneity using computed tomography *PhD Thesis* University of Toronto

Sontag M R and Cunningham J R 1977 Corrections to absorbed dose calculations for tissue inhomogeneities *Med. Phys.* **4** 431–6

——1978a Clinical application of a CT based treatment planning system *Comput. Tomogr.* **2** 117–30

——1978b The equivalent tissue-air ratio method for making absorbed dose calculations in a heterogeneous medium *Radiology* **129** 787–94

Sontag M R and Ray S K 1995 Determination of differential scatter-air ratios (dSAR) for three-dimensional scatter integration *Med. Phys.* **22** 775–80

Sterling T D, Perry H and Katz L 1964 Automation of radiation treatment planning IV. Derivation of a mathematical expression for the per cent depth dose surface of cobalt 60 beams and visualisation of multiple field dose distributions *Br. J. Radiol.* **37** 544–50

Storchi P and van Gasteren J J M 1996 A table of phantom scatter factors of photon beams as a function of the quality index and field size *Phys. Med. Biol.* **41** 563–71

Storchi P and Woudstra E 1996 Calculation of the absorbed dose distribution due to irregularly shaped photon beams using pencil beam kernels derived form basic beam data *Phys. Med. Biol.* **41** 637–56

Storchi P, Woudstra E, Verlinde P, Johansson K-A and Samuelsson A 1998 Calculation of absorbed dose distributions from dynamic wedges *Phys. Med. Biol.* **43** 1497–506

Sundblom I 1965 Dose planning for irradiation of thorax with ^{60}Co in fixed-beam teletherapy *Acta Radiol.* **3** 342–52

Svensson H and Brahme A 1986 Recent advances in electron and photon dosimetry *Radiation Dosimetry, Physical and Biological Aspects* ed C G Orton (New York: Plenum) pp 87–170

Svensson R and Brahme A 1996 Effective source size, yield and beam profile from multi-layered bremsstrahlung targets *Phys. Med. Biol.* **41** 1353–79

Svensson R, Källman P and Brahme A 1994 An analytical solution for the dynamic control of multileaf collimators *Phys. Med. Biol.* **39** 37–61

Swindell W, Simpson R G, Oleson J R, Chen C-T and Grubbs F A 1983 Computed tomography with a linear accelerator with radiotherapy applications *Med. Phys.* **10** 416–20

Tailor R C, Tello V M, Schroy C B, Vossler M and Hanson W F 1998 A generic off-axis energy correction for linac photon beam dosimetry *Med. Phys.* **25** 662–7

Tatcher M and Bjärngard B E 1993 Equivalent squares of irregular photon fields *Med. Phys.* **20** 1229–32

Thomas S J 1991 A modified power-law formula for inhomogeneity corrections in beams of high-energy x-rays *Med. Phys.* **18** 719–23

Treuer H, Boesecke R, Hartmann G H, Schlegel W and Lorenz W J 1987 Dosimetrische Bestimmung der Primärfluenz und der Fokusgrösse eines 15-MeV-Linearbeschleunigers. *Med. Physik* 375–80

Ulmer W 1982 On the application of stochastic partition functions for the computation of lateral profiles and depth doses in radiotherapy *Strahlentherapie* **158** 305–11

Uwamino Y, Nakamura T, Ohkubo T and Hara A 1986 Measurement and calculation of neutron leakage from a medical electron accelerator *Med. Phys.* **13** 374–84

van de Geijn J 1987 The extended net fractional depth dose: correction for inhomogeneities, including effects of electron transport in photon beam dose calculation *Med. Phys.* **14** 84–92

Van Dyk J, Barnett R B, Cygler J E and Shragge P C 1993 Commissioning and quality assurance of treatment planning computers *Int. J. Radiat. Oncol. Biol. Phys.* **26** 261–73

van Gasteren J J M, Heukelom S, van Kleffens H J, van der Laarse R, Venselaar J L M and Westermann C F 1991 The determination of phantom and collimator scatter components of the output of megavoltage photon beams: measurement of the collimator scatter part with a beam-coaxial narrow cylindrical phantom. *Radiother. Oncol.* **20** 250–7

van Santvoort J P C and Heijmen B J M 1996 Dynamic multileaf collimation without 'tongue-and-groove' underdosage effects *Phys. Med. Biol.* **41** 2091–105

van't Veld A A 1997 Analysis of accuracy in dose and position in calculations of a treatment planning system for blocked photon fields *Radiother. Oncol.* **45** 245–51

van't Veld A A and Bruinvis I A 1995 Influence of shape on the accuracy of grid-based volume computations *Med. Phys.* **22** 1377–85

Wagner J C, Redmond II E L, Palmtag S P and Hendricks J S 1994 MCNP: multigroup/adjoint capabilities *Report* LA-12704 (Los Alamos National Laboratory)

Wang L, Chui C-S and Lovelock M 1998 A patient-specific Monte Carlo dose-calculation method for photons beams *Med. Phys.* **25** 867–78

Wang L, Zhu Y and Jette D 1995 Analytic modeling of the primary x-ray dose deposition kernels (abstract) *Med. Phys.* **22** 977

Webb S 1993 *The Physics of Three-Dimensional Radiation Therapy: Conformal Radiotherapy, Radiosurgery and Treatment Planning (Medical Science Series)* (Bristol: Institute of Physics Publishing)

——1997 *The Physics of Conformal Radiotherapy Advances in Technology (Medical Science Series)* (Bristol: Institute of Physics Publishing)

——1998a Configuration options for intensity-modulated radiation therapy using multiple static fields shaped by a multileaf collimator *Phys. Med. Biol.* **43** 241–60

——1998b Configuration options for intensity-modulated radiation therapy using multiple static fields shaped by a multileaf collimator. II: Constraints and limitations on 2D modulation. *Phys. Med. Biol.* **43** 1481–95

Webb S and Fox R A 1980 Verification by Monte Carlo methods of a power law tissue-air ratio algorithm for inhomogeneity corrections in photon beam dose calculations *Phys. Med. Biol.* **25** 225–40

Weber L, Ahnesjö A, Nilsson P, Saxner M and Knöös T 1996 Verification and implementation of dynamic wedge calculations in a treatment planning system based on a dose-to-energy-fluence formalism *Med. Phys.* **23** 307–16

Weber L, Nilsson P and Ahnesjö A 1997 Build-up cap materials for mesurement of photon head-scatter factors *Phys. Med. Biol.* **42** 1875–86

Wong E, Van Dyk J and Zhu Y 1997 Lateral electron transport in FFT photon dose calculations *Med. Phys.* **24** 1992–2000

Wong E, Zhu Y and Van Dyk J 1996 Theoretical developments on fast Fourier transform convolution dose calculations in inhomogeneous media *Med. Phys.* **23** 1511–21

Wong J W and Henkelman R M 1982 Reconsideration of the power-law (Batho) equation for inhomogeneity corrections *Med. Phys.* **9** 521–30

——1983 A new approach to CT pixel-based photon dose calculations in heterogeneous media *Med. Phys.* **10** 199–208

Wong J W, Henkelman R M, Andrew J W, Van Dyk J and Johns H E 1981a Effect of small inhomogeneities on dose in a cobalt-60 beam *Med. Phys.* **8** 783–91

Wong J W, Henkelman R M, Fenster A and Johns H E 1981b Second scatter contribution to dose in a cobalt-60 beam *Med. Phys.* **8** 775–82

Wong J W and Purdy J A 1987 Basis of recent methods of photon dose calculations *9th Int. Conf. on The Use of Computers in Radiation Therapy (Scheveningen, The Netherlands)* ed I A D Bruinvis, P H van der Giessen, H J van Kleffens and F H Wittkampher (Amsterdam: Elsevier) pp 319–22

——1990 On methods of inhomogeneity corrections for photon transport. *Med. Phys.* **17** 807–14

——1992 Review of methods of inhomogeneity corrections *Advances in Radiation Oncology Physics: Dosimetry, Treatment Planning and Brachytherapy* (New York: American Institute of Physics) pp 887–99

Wong J W, Slessinger E D, Hermes R E, Offutt C J, Roy T and Vannier M W 1990 Portal dose images I: Quantitative treatment plan verification *Int. J. Radiat. Oncol. Biol. Phys.* **18** 1455–63

Wong J W, Slessinger E D, Rosenberger F U, Krippner K and Purdy J A 1984 The delta volume method for 3-dimensional photon dose calculations *Int. Conf. on the Use of Computers in Radiation Therapy, VIII ICCR (Toronto, Canada)* (Silver Spring, MD: IEEE Computer Society Press) pp 26–30

Woo M K 1994 Analysis of photon beam exit dose using photon point kernels *Phys. Med. Biol.* **39** 687–702

Woo M K and Cunningham J R 1990 The validity of the density scaling method in primary electron transport for photon and electron beams *Med. Phys.* **17** 187–94

Woo M K, Cunningham J R and Jezioranski J J 1990 Extending the concept of primary and scatter separation to the condition of electronic disequilibrium *Med. Phys.* **17** 588–95

Worthley B 1966 Equivalent square of rectangular fields (letter) *Br. J. Radiol.* **39** 559

Wrede D E 1972 Central axis tissue–air ratios as a function of area-perimeter at depth and their applicability to irregularly shaped fields *Phys. Med. Biol.* **17** 548–54

Yan D, Ziaja E, Jaffray D, Wong J, Brabbins D, Vicini F and Martinez A 1998 The use of adaptive radiation therapy to reduce setup error: a prospective clinical study *Int. J. Radiat. Oncol. Biol. Phys.* **41** 715–20

Ying X, Geer L Y and Wong J W 1990 Portal dose images II: Patient dose estimation *Int. J. Radiat. Oncol. Biol. Phys.* **18** 1465–75

Yu C X, Mackie T R and Wong J W 1995 Photon dose calculation incorporating explicit electron transport *Med. Phys.* **22** 1157–65

Yu C X and Wong J W 1993 Implementation of the ETAR method for 3D inhomogeneity correction using FFT *Med. Phys.* **20** 627–32

Yu M K, Sloboda R S and Mansour F 1996 Measurement of photon beam backscatter from collimators to the beam monitor chamber using target-current-pulse-counting and telescope techniques *Phys. Med. Biol.* **41** 1107–17

Yu M K, Sloboda R S and Murray B 1997 Linear accelerator photon beam quality at off-axis points *Med. Phys.* **24** 233–9

Zefkili S, Kappas C and Rosenwald J C 1994 On-axis and off-axis primary dose component in high energy photon beams *Med. Phys.* **21** 799–808

Zhu T C and Bjärngard B E 1995 The fraction of photons undergoing head scatter in x-ray beams *Phys. Med. Biol.* **40** 1127–34

Zhu T C and Palta J R 1998 Electron contamination in 8 and 18 MV photon beams *Med. Phys.* **25** 12–19

Zhu Y and Boyer A 1990 X-ray dose computations in heterogeneous media using 3-dimensional FFT convolution *Phys. Med. Biol.* **35** 351–68

Zhu Y and Van Dyk J 1995 Accuracy requirements of the primary x-ray spectrum in dose calculations using FFT convolution techniques *Med. Phys.* **22** 421–6

Report of the AAPM Task Group No. 105: Issues associated with clinical implementation of Monte Carlo-based photon and electron external beam treatment planning

Indrin J. Chetty[a)]
University of Michigan, Ann Arbor, Michigan 48109
and University of Nebraska Medical Center, Omaha, Nebraska 68198-7521

Bruce Curran
University of Michigan, Ann Arbor, Michigan 48109

Joanna E. Cygler
Ottawa Hospital Regional Cancer Center, Ottawa, Ontario K1H 1C4, Canada

John J. DeMarco
University of California, Los Angeles, Callifornia 90095

Gary Ezzell
Mayo Clinic Scottsdale, Scottsdale, Arizona 85259

Bruce A. Faddegon
University of California, San Francisco, California 94143

Iwan Kawrakow
National Research Council of Canada, Ottawa, Ontario K1A 0R6, Canada

Paul J. Keall
Stanford University Cancer Center, Stanford, California 94305-5847

Helen Liu
University of Texas MD Anderson Cancer Center, Houston, Texas 77030

C.-M. Charlie Ma
Fox Chase Cancer Center, Philadelphia, Pennsylvania 19111

D. W. O. Rogers
Carleton University, Ottawa, Ontario K1S 5B6, Canada

Jan Seuntjens
McGill University, Montreal, Quebec H3G 1A4, Canada

Daryoush Sheikh-Bagheri
The Regional Cancer Center, Erie, Pennsylvania 16505

Jeffrey V. Siebers
Virginia Commonwealth University, Richmond, Virginia 23298

(Received 11 April 2006; revised 11 July 2007; accepted for publication 18 September 2007; published 27 November 2007)

The Monte Carlo (MC) method has been shown through many research studies to calculate accurate dose distributions for clinical radiotherapy, particularly in heterogeneous patient tissues where the effects of electron transport cannot be accurately handled with conventional, deterministic dose algorithms. Despite its proven accuracy and the potential for improved dose distributions to influence treatment outcomes, the long calculation times previously associated with MC simulation rendered this method impractical for routine clinical treatment planning. However, the development of faster codes optimized for radiotherapy calculations and improvements in computer processor technology have substantially reduced calculation times to, in some instances, within minutes on a single processor. These advances have motivated several major treatment planning system vendors to embark upon the path of MC techniques. Several commercial vendors have already released or are currently in the process of releasing MC algorithms for photon and/or electron beam treatment planning. Consequently, the accessibility and use of MC treatment planning algorithms may well become widespread in the radiotherapy community. With MC simulation, dose is computed stochastically using first principles; this method is therefore quite different from conventional dose algorithms. Issues such as statistical uncertainties, the use of variance reduction techniques, the

ability to account for geometric details in the accelerator treatment head simulation, and other features, are all unique components of a MC treatment planning algorithm. Successful implementation by the clinical physicist of such a system will require an understanding of the basic principles of MC techniques. The purpose of this report, while providing education and review on the use of MC simulation in radiotherapy planning, is to set out, for both users and developers, the salient issues associated with clinical implementation and experimental verification of MC dose algorithms. As the MC method is an emerging technology, this report is not meant to be prescriptive. Rather, it is intended as a preliminary report to review the tenets of the MC method and to provide the framework upon which to build a comprehensive program for commissioning and routine quality assurance of MC-based treatment planning systems. © *2007 American Association of Physicists in Medicine.* [DOI: 10.1118/1.2795842]

Key words: Monte Carlo dose calculation, clinical treatment planning, experimental verification

TABLE OF CONTENTS

I. INTRODUCTION

I.A. Motivation

The accuracy of dose calculations is crucial to the quality of treatment planning and consequently to the doses delivered to patients undergoing radiation therapy.[1] Among other fac-

tors, dose calculations form an integral component in optimizing the therapeutic gain, i.e., maximizing the dose to the tumor for a given normal-tissue dose, for patients treated with radiation. Although the clinical benefit of more accurate dose distributions (i.e., how the improved dose distributions will affect tumor recurrence, i.e., local control, and normal tissue complications) has not been adequately quantified and requires further investigation, evidence exists that dose differences on the order of 7% are clinically detectable.[2] Moreover, several studies have shown that 5% changes in dose can result in 10%–20% changes in tumor control probability (TCP) or up to 20–30% changes in normal tissue complication probabilities (NCTP) if the prescribed dose falls along the steepest region of the dose-effect curves.[3–5] Readers interested in further understanding the need for heterogeneity corrections, among other topics related to dose calculations, are encouraged to read the AAPM Report No. 85,[1] where a comprehensive review of tissue heterogeneity corrections for megavoltage photon beams is provided.

In this report we focus our attention on the Monte Carlo (MC) method, a dose calculation algorithm known to be very accurate when used properly for treatment planning in heterogeneous patient tissues. The issue of lengthy calculation times has traditionally led to the MC method being viewed in the medical physics community as a clinically unfeasible approach. However, the development of MC codes optimized for radiotherapy calculations as well as the availability of much faster and affordable computers, have substantially reduced processing times. These significant advances have led to the clinical use of MC algorithms at some treatment centers and the promised availability of MC photon/electron planning modules among several commercial treatment planning vendors.

In light of the above considerations, MC treatment planning is quickly becoming a reality. An introductory report for the medical physics community on the understanding, implementation, testing, and use of MC algorithms is therefore warranted.

I.B. Objectives for the report

We intend this document to be a preliminary report with the following objectives: (a) to provide an educational review of the physics of the MC method and how it is applied in external beam radiotherapy dose calculations, (b) to describe the role of the MC method in external beam radiotherapy treatment planning process: from the interaction of electrons in the target of the linear accelerator to the deposition of dose in the patient tissues, (c) to describe the issues associated with MC dose calculation within the patient-specific geometry, (d) to discuss the issues associated with experimental verification of MC algorithms, and (e) to discuss the clinical implications of MC calculated dose distributions.

We expect that areas of concern outlined in this report will be further investigated and that more detailed reports providing recommendations on the major issues will be forthcoming.

I.C. Organization of the report

Following the introductory section (Sec. I) we begin in Sec. II with a review of the MC method as it applies to photon and electron transport. We include in this section an overview of MC simulation from the accelerator treatment head to the patient as well as a discussion of variance reduction techniques and efficiency-enhancing methods integral to MC calculations in radiotherapy. Section III begins with a review of the major MC codes being used in clinical application and is followed by detailed discussions on accelerator treatment head modeling and patient-specific treatment planning. This section concludes with the topic of experimental verification, in which guidance is provided on the types of tests needed to verify the accuracy of MC dose calculation algorithms. In Sec. IV we provide a review of recent studies demonstrating the potential clinical impact of MC dose calculations in comparison with conventional algorithms. Finally, we conclude in Sec. V with a summary of the recommendations from this task group report.

II. THE MONTE CARLO METHOD IN RADIOTHERAPY DOSE CALCULATIONS

II.A. Definition of the MC method and historical background

Most generally, the MC technique is a statistical method for performing numerical integrations. MC simulations are employed in many areas of science and technology. Although a method based on random sampling was discussed as early as 1777 by Buffon,[6] the MC technique as we know it today was first developed and named at the end of the second world war. The motivation was to apply MC techniques to radiation transport, specifically for nuclear weapons.[7] The driving forces for the initial idea appear to have been Stanislaw Ulam and John von Neumann who saw the development of ENIAC, the first electronic computer, as an ideal opportunity to develop new applications of statistical sampling. The developments of MC techniques and computers have been closely intertwined ever since, with an exponential increase of the application of MC simulations since digital computers became widely available in the 1950s and 1960s.

Modeling of particle transport problems is ideally suited for the use of MC methods and has been described by Rogers and Bielajew as follows: "The Monte Carlo technique for the simulation of the transport of electrons and photons though bulk media consists of using knowledge of the probability distributions governing the individual interactions of electrons and photons in materials to simulate the random trajectories of individual particles. One keeps track of physical quantities of interest for a large number of histories to provide the required information about the average quantities."[8] As a technique for calculating dose in a patient the underlying physical basis is much simpler in concept than analytic algorithms because the MC method consists of a straightforward simulation of reality and does not involve complex approximations nor models of dose deposition, but only a knowledge of the physics of the various interactions which

have been well understood for over 50 years in most cases. While some of these interactions may be complex to simulate in detail, the basic ideas of each interaction, e.g., an electron giving off a bremsstrahlung photon, are well understood by medical physicists and, hence, the overall process is easy to comprehend.

Although MC was used in several particle physics applications to simulate electron-photon showers in the 1950s, the seminal paper in the field was that of Berger in 1963,[9] in which he described the condensed history technique for electron transport. This technique is the basis of all modern electron-photon transport MC codes relevant to medical physics. The ETRAN code, based on these ideas[10] was developed by Berger and Seltzer and now forms the basis of electron transport in the MCNP code.[11] The release of the EGS4 MC code system in 1985[12] served as a catalyst for the application of the MC method in radiotherapy calculations of dose and dosimeter response. The work of Petti *et al.*,[13] Mohan *et al.*,[14] and Udale[15] being early examples of the use of the EGS MC code system[12,16,17] to simulate medical linear accelerators. Even without the direct use of MC simulations, the MC method already plays a significant role in radiotherapy treatment planning since the energy deposition kernels used in convolution/superposition algorithms have been calculated using MC techniques. Linear accelerator calibration protocols (e.g., AAPM's TG-51)[18] use factors derived from MC simulations. MC-based calculations are also used in the design of treatment head components.[19,20]

Although it has only recently become practical, for over two decades the application of MC techniques to radiation treatment planning has been quite clear.[21,22] The widely used BEAM code system[23] is a pair of EGS4 (now EGSnrc)[17] user codes for simulating radiation transport in accelerators and in patients represented by CT data sets. These relatively easy to use tools have sparked intense research in MC-based radiotherapy treatment planning and have lead to two comprehensive reviews of accelerator simulations by Ma and Jiang[24] and Verhaegen and Seuntjens.[25] Kawrakow and Fippel, among others, have provided the breakthroughs which have made clinical treatment planning feasible, as discussed in Sec. III A. The fast MC codes being developed commercially are almost all based on the results of this collaboration. As this report is being written, the first commercial MC systems have already been introduced into routine clinical treatment planning for electrons[26] and photons.[27]

II.B. Monte Carlo simulation of electron and photon transport

The following material represents a very brief introduction into the MC simulation of electron and photon transport. For more details the reader is referred to the reviews available in the literature.[8,28–31] Another source for detailed information is the documentation accompanying some of the general purpose codes, for instance, the EGSnrc,[17] MCNP,[32] and GEANT4 (Ref. 33) manuals, and the PENELOPE paper.[34]

In the energy range of interest for external beam radiotherapy (megavoltage range), photons interact with surrounding matter via four main processes: incoherent (Compton) scattering with atomic electrons, pair production in the nuclear or electron electromagnetic field, photoelectric absorption, and coherent (Rayleigh) scattering. The first three collision types transfer energy from the photon radiation field to electrons or positrons. In most cases Compton scattering is the dominant interaction, although pair production becomes increasingly important with increasing energy, and may even dominate at higher energies in high-Z components of the treatment head of medical linear accelerators.

When electrons traverse matter, they undergo a large number of elastic interactions and lose energy by two main processes: inelastic collisions with atoms and molecules and radiative interactions. Inelastic collisions result in excitations and ionizations. Ionizations lead to secondary electrons, sometimes referred to as "δ particles". Radiative energy losses, which occur in the form of bremsstrahlung and positron annihilation, transfer energy back to photons and lead to the coupling of the electron and photon radiation fields. One therefore speaks of coupled electron-photon showers.

The electron-photon macroscopic radiation field can be described mathematically by a coupled set of integrodifferential transport equations. These transport equations are prohibitively complicated thereby excluding an analytical treatment except under severe approximations. The MC technique is a solution method that can be applied for any energy range and underlying geometry and material composition.

A solution of the transport problem of particles in matter, which is exact within the existing knowledge of the elementary collision processes, can be obtained by an analog MC simulation. In an analog simulation all particle interactions with surrounding atoms and molecules are explicitly simulated, including those of secondary particles created in the collisions. An analog MC technique is therefore a faithful simulation of physical reality on a digital computer: particles (photons for example) are "born" according to distributions describing the source, they travel a certain distance, determined by a probability distribution, to the site of a collision, and scatter into another energy and/or direction state, possibly creating additional particles. These photons eventually "die" as a result of pair production or photoelectric events or when they Compton scatter to energies below a predetermined low-energy photon cutoff, often called PCUT. Analog simulations, often referred to as "event-by-event" or "interaction-by-interaction" techniques, are typically used for the transport of neutral particles. The analog simulation of charged particle transport is not practical, due to the large number of interactions they undergo until locally absorbed as the energy of the charged particle falls below the predetermined low-energy limit for tracking charged particles, often call ECUT. All general purpose MC codes therefore employ condensed history schemes for charged particle transport, discussed in more detail in Sec. II B 2.

Within a MC simulation, quantities of interest can be computed by averaging over a given set of particle showers (also referred to as "histories," "cases," "trajectories," or "tracks"). One can calculate both observable (measurable)

quantities, such as dose or a particle spectrum, and quantities that cannot easily be measured such as the fraction of particles originating from a certain component of the treatment head, the dose fraction due to scattered photons, etc. Although there are techniques for scoring quantities at a point when using Monte Carlo techniques, in treatment planning applications, it is usual to score quantities (dose mainly) averaged over some finite volume or voxel. As the voxel size is increased, for a given statistical uncertainty the total calculation time will decrease, but the spatial resolution is reduced (see Sec. III D 6 for more discussion). Another important aspect of MC calculations is the presence of statistical uncertainties due to the statistical nature of the method, which is discussed in more detail in Sec. III D 1.

II.B.1. Analog simulations

An analog simulation of particle transport consists of four main steps:

(1) Select the distance to the next interaction.
(2) Transport the particle to the interaction site taking into account geometry constraints.
(3) Select the interaction type.
(4) Simulate the selected interaction.

Steps 1–4 are repeated until the original particle and all secondary particles leave the geometry or are locally absorbed. A particle is considered to be locally absorbed when its energy falls below a specified threshold energy.

Step 1 is based on the probability, $p(r)dr$, that a particle interacts in an interval dr at a distance r from its initial position

$$p(r)dr = e^{-\mu r}\mu dr, \tag{1}$$

where μ is the linear attenuation coefficient (number of interactions per unit length). A random distance r distributed according to $p(r)$ can be sampled using the so-called inverse-transform method, which equates the cumulative probability of $p(r)$ with a random number ξ distributed uniformly between zero and unity

$$\int_0^r p(r')dr' = \xi \Rightarrow r = -\frac{\ln(1-\xi)}{\mu}. \tag{2}$$

Step 2 involves basic ray tracing, which requires a geometry model that can provide the medium and mass density of a region together with a computation of the distance to the next geometry boundary along the particle trajectory.

Step 3 is similar to step 1 except that now the probability distribution function is discrete, i.e., it involves a fixed number of final states i, corresponding to an interaction of type i. Suppose that the cross section for interaction of type i is denoted by σ_i and the total cross section by $\sigma = \Sigma\sigma_i$. A direct application of the inverse-transform method for n interaction types yields interaction *1* if $\xi \leq \sigma_1/\sigma$, else interaction *2* if $\xi \leq (\sigma_1+\sigma_2)/\sigma$, else interaction n, if $\xi \leq (\sigma_1+\sigma_2+\cdots+\sigma_n)/\sigma$.

Perhaps the most difficult part is step 4, where one must sample energy/direction changes from the differential cross section of the selected process. The manuals of the popular general purpose codes[17,32–34] provide details of the methods employed for the relevant photon interactions.

Based on the above discussion it should be clear that an analog MC simulation is conceptually quite straightforward.

II.B.2. Condensed history simulations

The condensed history technique was first described comprehensively in the pioneering work by Berger.[9] It is based on the observation that the vast majority of electron interactions lead to very small changes in the electron energy and/or direction. Many such "small-effect" interactions can therefore be grouped into relatively few condensed history "steps" and their cumulative effect taken into account by sampling energy, direction, and position changes from appropriate distributions of grouped single interactions, e.g., multiple scattering, stopping power, etc. Berger defined two main classes of condensed history implementations. In a class I scheme all collisions are subject to grouping. The effect of secondary particle creation above specified threshold energies are taken into account after the fact (i.e., independently of the energy actually lost during the step) by setting up and transporting the appropriate number of secondary particles. In this way the correlation between large energy losses and secondary particle creation is lost. In a class II scheme interactions are divided into "hard" (sometimes also referred to as "catastrophic") and "soft" collisions. Soft collisions are subject to grouping as in a class I scheme; hard collisions are explicitly simulated in an analog manner.

A class II scheme can be described with the same four basic steps that make up an analog simulation. The two main differences are that only hard collisions are included and that step 2 is much more difficult because the particles do not move on straight lines and because it involves the selection of energy, direction, and position changes from multiple scattering distributions. It is also frequently necessary to divide the distance between catastrophic interactions into shorter condensed history steps to guarantee the accuracy of the simulation. As in an analog simulation there is a transport threshold energy. Particle transport thresholds are often the same as the particle production thresholds dividing hard and soft collisions, but this is not a necessary condition.

A class II MC simulation is illustrated in Fig. 1. The upper portion shows a complete electron track including secondary electrons and photons (shown with dashed lines and not including their interactions) with energies above the hard collision thresholds. The lower portion is a magnified view of the shaded box. The actual curved path has been simulated using four condensed history steps. The filled circles and arrows show the positions and directions at the beginning of the steps. The shaded area around the electron track indicates the region where the energy of subthreshold secondary particles is in reality deposited. If this volume is small compared to the calculation resolution (i.e., voxel size in the case of radiotherapy calculations), energy deposition can be considered local and modeled using a restricted stopping power along the electron track. Note that the initial and final posi-

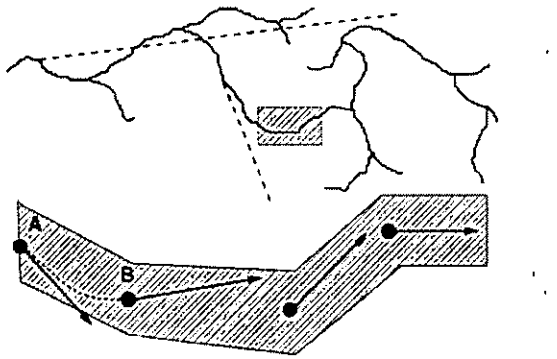


FIG. 1. Illustration of a class II condensed history scheme for electron transport. The upper portion shows a complete electron track including secondary electrons and photons (shown with dashed lines and not including their interactions) with energies above the hard collision thresholds. The lower portion is a magnified view of the shaded box.

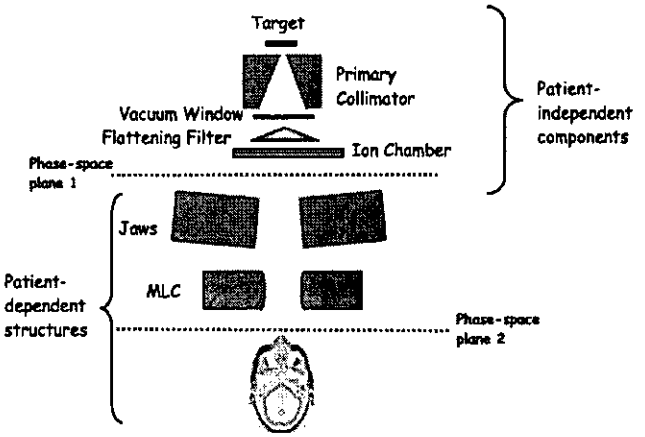


FIG. 2. Illustration of the components of a typical Varian linear accelerator treatment head in photon beam mode. Phase space planes for simulating patient-dependent and patient-independent structures are also represented. For other manufacturers, component structures (such as the jaws, MLC, etc.) may be in different locations, thereby potentially requiring a change in the placement of the phase space scoring planes.

tions of a step are not connected in the figure to highlight the fact that a condensed history implementation does not provide information on how the particle goes from *A* to *B* (the curved dashed line connecting *A* and *B* is a more realistic representation of the trajectory than a straight line from *A* to *B*). This becomes important when the scoring grid is not the same as the underlying geometry grid or when a single condensed history step traverses more than one geometrical region as in some of the fast MC codes specialized for use in radiotherapy.[35–39] This consideration highlights another important aspect of a condensed history simulation, namely the way the transport is performed in the vicinity of or across boundaries between different regions. The EGSnrc code, for instance, utilizes single scattering (i.e., analog) simulation within a certain perpendicular distance from an interface.[40] Although this approach is necessary for accurate simulations of certain types of geometries, it is generally not needed for typical radiotherapy calculations.

Although the condensed history technique makes use of practical MC simulations possible, it introduces the step size as an artificial parameter. Dependencies of the calculated results on the step size have become known as step-size artifacts.[41] Step-size artifacts were a major factor in the early years of most general purpose MC codes. Due to significant theoretical developments in the nineties the condensed history technique is now well understood.[42,43] This has lead to the development of high accuracy condensed history implementations[40,44] and faster MC codes that can compute dose distributions with accuracy comparable to traditional MC packages in a small fraction of the time.[35–39,45]

With charged particle transport one stops tracking the particle's movement at some low-energy cutoff and the choice of the cutoff can affect the calculation in two important ways. The higher the value of the cutoff, the faster the calculation; this can improve the calculation speed significantly. On the other hand, unless great care is taken, stopping at too high a cutoff energy can distort the dose distribution since the "stopped" charged particle might have deposited energy some distance from where its trajectory was terminated. Thus care must be taken in selecting an energy cutoff.

II.C. Overview of Monte Carlo-based radiotherapy dose calculations

It is possible to carry out a single MC simulation in which one starts with the electron exiting from the accelerator structure, follows it and its descendants (e.g., bremsstrahlung photons, knock-on electrons) through the fixed elements of the head (targets or scattering foils, primary collimators, monitor chambers, flattening filters, etc.), the various beam shaping devices which are patient specific (jaws, multileaf collimator (MLC), applicators, cutouts, wedges, compensators), and finally the patient as specified by a CT or some other data set. This tracking of the initial particle and all of its descendants is referred to as a history. As discussed further in Sec. III D 1.1, it is important to include all particles associated with a single initiating electron as part of the same history.

Due to significant improvements in the efficiency of photon beam treatment head simulations,[46] the speed of the complete simulation is such that it is feasible to consider performing the entire calculation for each patient.[47] However, there have been a variety of strategies for dividing such calculations into several steps. The first step, transporting particles through the patient-independent elements, can be inefficient without the use of advanced variance reduction techniques (see Sec. II D). This is especially true for photon beams, since many bremsstrahlung photons generated in the target will strike the primary collimator and not contribute to the beam reaching the patient. One approach to improve the simulation efficiency is to first perform the simulation of the patient-independent structures and to store what is called a phase-space file at a plane just below the fixed elements of the accelerator head (see phase-space plane 1 in Fig. 2). The phase-space file contains phase-space parameters for all particles as they cross the scoring plane. The phase-space parameters consist of the energy, position, direction, charge, and possibly other information such as the region/s of cre-

ation or interaction of particles. The advantage of this approach is that this part of the calculation can be reused as often as necessary. Particles are then transported through the patient specific collimation system and are either stored in another phase-space file at the base of the accelerator (see phase-space plane 2 in Fig. 2) or tracked through the patient in the same simulation. Storing a second phase-space may be more efficient when open fields (e.g., 10×10 cm^2 fields) are used for treatment, however, more commonly, when MLCs are used for beam shaping, the latter approach is likely to be more efficient.

As we will discuss in Sec. III C, the entire phase-space data for the accelerator may also be generated using beam modeling (virtual source model) approaches which do not require direct MC simulation of the accelerator for each treatment. One class of virtual source models is based on characterizing the results of a MC simulation of the accelerator head and another class is based solely on measured beam data such as depth-dose curves, profiles and output ratios. In either case, the patient-dependent components (e.g., the MLC) are simulated using either explicit transport methods or approximate transport methods before detailed transport in the patient.

As with conventional planning methods, experimental verification forms an integral part of the clinical implementation of a MC dose calculation algorithm. MC algorithms will benefit from similar experimental verification procedures as conventional systems, and as such should follow the commissioning procedures detailed in the AAPM TG-53 report[48] and other relevant publications.[49,50] Experimental verification in more complex fields and/or heterogeneous geometries is useful to verify the expected improved accuracy associated with MC calculations in these situations.

II.D. Variance reduction techniques and efficiency enhancing methods

The efficiency, epsilon (ε), of a MC calculation is defined as: $\varepsilon=1/s^2T$, where s^2 is an estimate of the true variance (σ^2) of the quantity of interest and T is the CPU time required to obtain this variance. Since both Ns^2 and T/N are approximately constant, the efficiency is roughly independent of N, the number of histories simulated. There are two ways to improve the efficiency of a given calculation: either decrease s^2 for a given T or decrease T for a given N while not changing the variance. Techniques which improve the efficiency by changing the variance for a given N while not biasing the result (i.e., not changing the expectation value which is the value expected in an infinitely long run) are called variance reduction techniques. Variance reduction techniques often increase the time to simulate a single history and are only useful if the overall efficiency is improved. A given technique may increase the efficiency for some quantities being scored and decrease it for others. In contrast to variance reduction techniques, there are a variety of ways to speed up a given calculation by making an approximation which may or may not affect the final result in a significant way.

Bremsstrahlung splitting and Russian roulette of secondary particles are widely used variance reduction techniques which are especially useful in simulating an accelerator treatment head.[23,46] In the various forms of bremsstrahlung splitting, each time an electron is about to produce a bremsstrahlung secondary, a large number of secondary photons with lower weights are set in motion, the number possibly depending on a variety of factors related to the likelihood of them being in the field. If the number of photons created is selected to minimize those that are not directed toward the patient plane, then there is a further saving in time. Russian roulette can be played whenever there is little interest in a particle resulting from a specific class of events. The low interest particles are eliminated with a given probability, but to ensure an unbiased result, the weights of the surviving particles are increased by the inverse of that probability. A common example is to play Russian roulette with secondary electrons created from photon interactions in treatment head structures. Another variance reduction method, photon forcing, may sometimes be used to enhance the production of electrons in the air downstream of the accelerator. In a photon forcing scheme, the parent photon is forced to interact in a given geometric region and the weights of the resulting particles are adjusted accordingly to maintain an unbiased result.

Range rejection and increasing the energy at which electron histories are terminated (energy cutoffs) are examples of methods which, when used correctly, improve efficiency by decreasing the time per history without significantly changing the results. In range rejection, an electron's history is terminated whenever its residual range is so short that it cannot escape from the current region or reach the region of interest. In most implementations this ignores the possible creation of bremsstrahlung photons while the electron loses energy which means this is an approximate technique. When applied to electrons below a certain energy threshold, this form of range rejection produces the same results in a reduced computing time.[23] It is also possible to implement range rejection in a manner which properly accounts for bremsstrahlung production and thus make it an unbiased variance reduction technique. By stopping tracking of electrons at a higher energy, efficiency can be improved, but this may have an effect on the dose distribution if too high a threshold is used. Playing Russian roulette with particles at energies below a relatively high transport cutoff or with range-rejected particles is a comparable variance reduction technique for reducing the simulation time. However, its implementation is typically more difficult and this has favored the use of range rejection and high transport cutoffs in situations where it is easy to demonstrate that the resulting error is sufficiently small.

There are other variance reduction and efficiency enhancing techniques which collectively have allowed substantial increases in the speed of a calculation. These methods include the reuse of particle tracks,[37] and other adaptations of particle track reuse, such as the simultaneous transport of particle sets (STOPS) approach of Kawrakow,[36] which is a variance reduction technique. For a more comprehensive re-

view of variance reduction and efficiency enhancing techniques, readers are referred to a chapter by Rogers and Bielajew[51] and the papers by Rogers *et al.*[23] and Kawrakow *et al.*[46,52] Other useful references on the influence of variance reduction methods in the context of phantom and patient treatment planning are provided in the articles by Ma *et al.*[53] and Kawrakow and Fippel.[45]

In summary, variance reduction techniques are an important requirement for the use of MC calculations in the clinical setting; without them, calculation times would still be too long for use in most situations. However, inappropriate use of a variance reduction technique can reduce calculation efficiency, thus increasing calculation time. In principle variance reduction techniques, implemented correctly, do not alter the physics and thereby produce unbiased results. Other efficiency improving techniques can significantly alter the accuracy of the calculation if applied inappropriately. Improper implementations can lead to unpredictable results in either case.

It is incumbent upon the medical physicist to understand, at a minimum, those techniques that the user can adjust in a clinical MC algorithm. In addition, tests must be done to show correct implementation of those techniques over the range of clinical situations. Vendors should provide adequate documentation for users to understand the techniques employed and how the implementation was validated.

III. MONTE CARLO SIMULATION OF RADIATION TRANSPORT IN ACCELERATORS AND PATIENTS

III.A. Review of current Monte Carlo codes

A large number of general purpose MC algorithms have been developed for simulating the transport of electrons and photons. Perhaps the most widely used of these in medical physics is the EGS code system.[12,16,17,40,44] There are several other comparable general purpose systems used in medical physics such as the ITS (Refs. 54 and 55) and MCNP systems[11,32] both of which have incorporated the electron transport algorithms from ETRAN (Refs. 10 and 56) which was developed at NIST by Berger and Seltzer following the condensed history techniques proposed by Berger.[9] Other newer general purpose systems include PENELOPE (Ref. 34) and GEANT4.[33] The EGS and ITS/ETRAN and MCNP systems are roughly of the same efficiency for calculations in very simple geometries when no variance reduction techniques are used, whereas the other systems tend to be considerably slower. An important special purpose code is the EGS user code, BEAM.[23,57–60] The BEAM code is optimized to simulate the treatment head of radiotherapy accelerators and includes a number of variance reduction techniques to enhance the efficiency of the simulation.[46] Comprehensive reviews of MC simulation of radiotherapy beams from linear accelerators are available elsewhere.[24,25]

While the accuracy of these general-purpose codes can be roughly the same as long as they are carefully used, these codes are considered too slow for routine treatment planning purposes. Several groups have published on the use of parallelization of MC techniques over multiple computers to provide more reasonable turn-around times for simulation in clinical research.[61–63] Specific to radiation therapy, there have been a variety of MC codes developed to improve the calculation efficiency, especially in the patient simulation. The PEREGRINE system (North American Scientific: Nomos Division) was developed at the Lawrence Livermore National Laboratory and has been benchmarked against measurements.[27] The PEREGRINE electron transport algorithm is a modified version of the EGS4 condensed history implementation. PEREGRINE uses the random hinge approach[64] for electron transport mechanics. Several efficiency enhancing and variance reduction techniques are implemented in PEREGRINE, including source particle reuse, range rejection, Russian roulette and photon splitting. Parallelizing the calculation on several computer processors is also implemented to reduce the overall dose calculation time. The system decouples the scoring zones from the transport geometry. Source modeling in PEREGRINE is achieved by performing a full MC simulation of the accelerator head using the BEAM code[23] and using the output to create a source model[65] from which source particles are regenerated above the patient-dependent beam modifiers. PEREGRINE uses several approximations when transporting the beam through the patient specific beam modifiers, followed by transport through a patient's CT data set.[66] The PEREGRINE system was the first MC algorithm to receive FDA 510-K approval and represents the first commercially available photon beam treatment planning system in the United States.

Several commercial MC implementations currently available or under development are based on the Voxel Monte Carlo (VMC) series of codes. The initial version[37] of VMC was only applicable to electron beams and involved several approximations in the modeling of the underlying interaction processes. Improved treatment of multiple elastic scattering[67] was incorporated in 1996, PENELOPE's random hinge method in 1997,[68] and all remaining approximations removed in 2000.[45] A photon transport algorithm was added in 1998,[35] which included precalculated interaction densities in each voxel similar to approaches developed previously.[69,70] The resulting code was named XVMC. In 1999, a series of advanced variance reduction techniques were developed and incorporated into XVMC which brought an additional factor of 5–9 increase in simulation speed. Treatment planning applications and experimental verification of VMC-based systems have been reported in several articles.[71–75] Separate versions of the VMC code were subsequently developed by Fippel (XVMC) (Refs. 45 and 76 and Kawrakow (VMC++)[36]. XVMC is being incorporated into the Monaco (CMS), PrecisePlan (Elekta), and *i*Plan (BrainLab) treatment planning systems. VMC++ includes additional refinements in the physics models, such as the exact Kawrakow–Bielajew multiple scattering formalism,[77] including relativistic spin effects,[17] and the STOPS method (mentioned in Sec. II D). VMC++ is the basis for the first commercial electron MC algorithm from Nucletron and is being incorporated into the Masterplan (Nucletron) and Eclipse (Varian) treatment planning systems for photon beam dose calculations. The VMC/XVMC/

VMC++ code systems have also been integrated into several MC-based research systems including those at the University of Tubingen, McGill University, and the Virginia Commonwealth University.

Another MC code that has reached commercial implementation is the Macro Monte Carlo (MMC) method[38,78] for electron beam treatment planning. MMC uses the MC technique, but is very different from the standard simulation of radiation transport. MMC uses a precalculated database from EGSnrc simulations of electron transport through small spheres of varying sizes and materials and follows a random walk through the CT phantom based on these precalculated values. The commercial implementation of MMC, eMC, (Eclipse, Varian) makes use of some precalculated accelerator-specific information; however, fluence intensities arising from the various subsources are fitted to the user's measured data.[79] More details on the performance of the eMC system for clinical electron dose calculations is available elsewhere.[80]

There are several institutions currently engaged in developing MC radiotherapy applications for clinical and/or research related purposes. MCDOSE (Ref. 53 and 81) is among the first of these types of systems. MCDOSE is based on EGS4 and includes fundamental changes in some aspects of the electron transport in order to improve speed. MCDOSE has been shown to give results very similar to EGS4.[53] It performs particle tracking through the beam modifiers in conjunction with the patient calculation and has built-in capability to handle various models of the incident beam. The speed ups have been obtained by using various techniques (bremsstrahlung splitting, photon forcing, track repetition,[37] and range rejection).

The MCV (Monte Carlo Vista) [82] code is used for clinical IMRT planning and verification[61] as well as for a variety of research related applications. MCV interfaces photon-electron MC dose algorithms to the Pinnacle (Philips Radiation Oncology Systems, Madison, WI) commercial planning system, and calculations are performed in a parallel environment using multiple Unix-based processors.[82] Treatment head simulation is accomplished using a modified version of the BEAM code, with calculations divided into two stages, based on the patient-dependent[83] and patient-independent component structures.[82] Patient and phantom calculations (within MCV) are completed using DOSXYZnrc,[58] VMC++ (described above), or MCVRTP, a C++ MC code developed by Philips that uses many of the algorithms of EGS4.[82] MCV utilizes variance reduction techniques inherent to the subcodes it uses, and achieves speed by use of multiple processors.[82]

Another major research code is the dose planning method[39] (DPM) code system developed initially for performing electron beam dose calculations in a voxelized geometry. DPM utilizes the Kawrakow–Bielajew multiple scattering formalism[77] and the random hinge approach for transport mechanics.[64] Particles do not stop at boundaries[39] as is the case with other fast codes. DPM has been integrated into the University of Michigan's in-house treatment planning system (UMPlan) and is currently being used for a variety of photon beam treatment planning studies.[84,85]

An MCNP (Monte Carlo *N*-particle)-based code, RT_MCNP (Ref. 86) has been in use for treatment planning research at UCLA. There have been a series of publications related to the use of RT_MCNP for a variety of applications, from radiosurgery to IMRT planning using a micromultileaf collimator.[86–91] Finally, treatment planning studies using the GEANT,[92–94] PENELOPE,[95] gamma electron positron transport system (GEPTS),[96,97] and ORANGE (Ref. 98) (MCNP-based) MC codes have also been reported.

In an effort to quantify the speed and accuracy for the phantom component of the calculations by the various MC codes being used for research and/or clinical planning purposes, Rogers and Mohan[99] proposed what came to be known as the ICCR benchmark. The tests and geometries for the ICCR benchmark comparisons were as follows:[99] (a) speed test: phantom of dimensions $30.5 \times 30.5 \times 30$ cm with $(5\ \text{mm})^3$ voxels filled either randomly with one of 4 materials (water, aluminum, lung, and graphite) or with water alone, 6 MV photons (spectrum) from a point source at 100 cm SSD and collimated to $10 \times 10\ \text{cm}^2$ at the phantom surface, (b) accuracy test: heterogeneous phantom as defined in (a) with $5 \times 5 \times 2$ mm voxels (2 mm along the depth axis), 18 MV photons (spectrum) from a point source at 100 cm SSD and collimated to $1.5 \times 1.5\ \text{cm}^2$ at the phantom surface. Beam spectra[268] were provided for these comparisons in order to standardize the beam model used in the dose calculations. Statistical uncertainties were to be reported as the relative uncertainty in the dose for voxels with a dose greater than some arbitrary lower limit, such as 50% of the maximum dose.[99] Results for the ICCR benchmark are summarized in Table I. Some timing results have been added recently. Timing values have been scaled to that on a single Intel P-IV 3.0 GHz processor.

The reader should be aware that the timing results reported in Table I are susceptible to large variations (on the order of at least 20%) due to variations in compilers, memory size, cache, etc.[99] Timing comparisons for more clinically relevant treatment plans are presented in Sec. III E 4. These results have been reported by medical physicists using commercial MC systems for treatment planning.

III.B. Accelerator treatment head simulation

III.B.1. Sensitivity of simulations to electron beam and other parameters

In general one does not know all the details of the clinical accelerator. For example, the characteristics of the incident electron beam are only known approximately. Knowledge of the sensitivity of MC simulation results to input parameters, such as the position, direction, and energy of the initial electron beam exiting the accelerator and to details of the geometry of the treatment head, is important. A sensitivity analysis is indispensable in determining which source and geometry parameters to adjust and by how much in order to improve agreement with user-specific measurements.

Factors influencing the characteristics of a photon beam are the energy, spatial, and angular distributions of the electrons incident on the target (or exiting the waveguide), and

 316

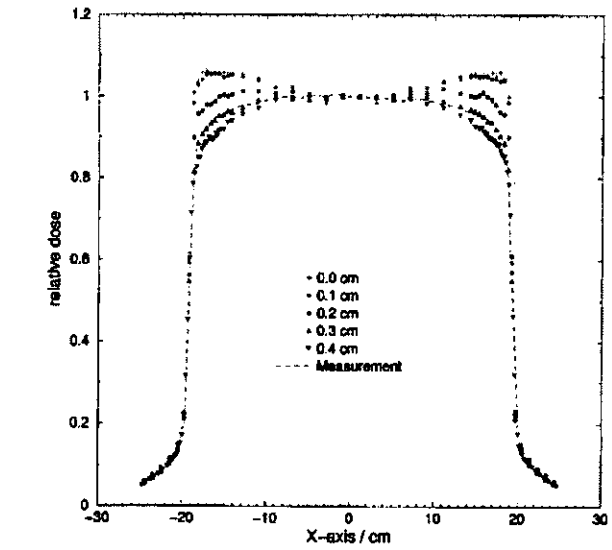
TABLE I. Summary of timing and accuracy results from the ICCR benchmark. Timing comparisons were performed using 6 MV photons, 10×10 cm^2 field size, and those for the accuracy test, using 18 MV photons and a 1.5×1.5 cm^2 field size, as detailed in the ICCR benchmark (Ref. 99). All times have been scaled to the time it would take running on a single, Pentium IV, 3 GHz processor. Readers should be aware that the timing results, as well as the method used to scale the times, are subject to large uncertainties due to differences in compilers, memory size, cache size, etc.

Monte Carlo code	Time estimate (min)	% mean difference relative to ESG4/PRESTA/DOSXYZ
ESG4/PRESTA/DOSXYZ	43	0, benchmark calculation
VMC++	0.9	±1
XVMC	1.1[a]	±1
MCDOSE (modified ESG4/PRESTA)	1.6	±1
MCV (modified ESG4/PRESTA)	22	±1
DPM (modified DPM)	7.3[b]	±1
MCNPX	60[c]	Maximum difference of 8% at Al/lung interface (on average ±1% agreement)
PEREGRINE	43[d]	±1
GEANT4 (4.6.1)	193[e]	±1 for homogeneous water and water/air interfaces

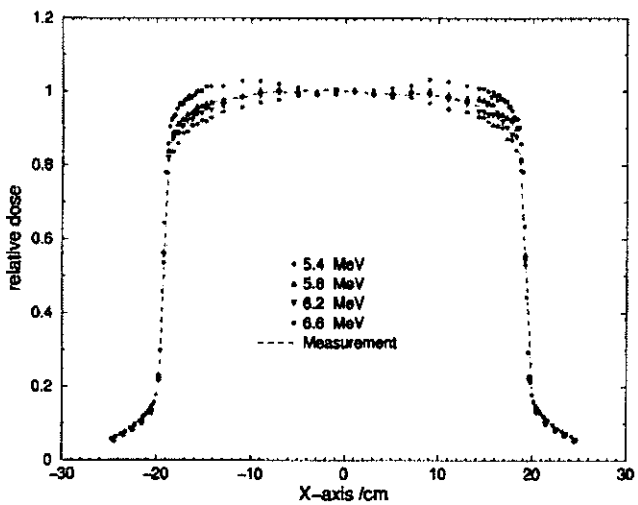
[a]Results not originally part of the ICCR benchmark study, from Ref. 76.
[b]Results not originally part of the ICCR benchmark study, reported independently by author I.J.C. for the modified version of DPM developed for clinical planning calculations.
[c]Timing results not originally part of the ICCR benchmark study, reported independently by author J.J.D. Calculations were performed using the *F8 (energy deposition) tally.
[d]Timing and accuracy results not originally part of the ICCR benchmark study, reported independently by author D.S.-B.
[e]Timing and accuracy results not originally part of the ICCR benchmark study. Estimated independently by author Seuntjens based on Poon and Verhaegen (Ref. 94) and ICCR type speed-test dose calculations (Ref. 270). Timing results are for the standard physics model; the low-energy and the PENELOPE model lead to a factor of 2 more CPU time. The accuracy result reported is derived from the interface perturbation studies in Poon and Verhaegen (Ref. 94), applies to 1.25 MeV monoenergetic photon beams and represents the difference with EGSnrc using PRESTA-II electron step algorithm and "exact" boundary crossing. For a water/Pb interface and 1.25 MeV photons, the maximum difference with EGSnrc increases to 6% (Ref. 94).

the dimensions, materials, and densities of all the components interacting with the beam (the target, primary collimator, flattening filter(s), monitor chamber, collimating devices, such as blocks or MLCs, and beam modifying devices such as wedges). Several investigators have reported on the sensitivity of megavoltage beam simulations to the electron beam striking the target and other treatment head parameters.[19,100–103] Faddegon *et al.*,[19] in simulating Siemens accelerators, showed that the key parameters are the mean energy and focal spot size of the electron beam incident on the exit window, the material composition and thickness profile of the exit window, target, flattening filter, primary collimator, and the position of the primary collimator relative to the target. Bieda *et al.*[101] showed that the accelerator simulation for 20 MeV (Varian-produced) electron beams was very sensitive to the distance between the scattering foils and, to a lesser extent, to the width of the shaped secondary scattering foil. Changes to the primary or secondary foil thickness were found to significantly alter the falloff and bremsstrahlung components of the depth-dose curve.[101] Sheikh-Bagheri and Rogers[102] performed calculations of "in-air" off-axis ratios and depth-dose curves and compared these with measurements to derive estimates for the parameters of the electron beam incident on the target, and to study the effects of some mechanical parameters, such as target width, primary collimator opening, flattening filter material, and density. Their study[102] included several different photon beam energies from accelerators produced by different manufacturers (Varian, Siemens, and Elekta). The electron beam radial intensity distribution was found to influence the off-axis ratios to a great extent. The greater the width of the electron-beam radial intensity distribution, the relatively more intense is the photon beam on the central axis.[102] Figure 3 shows the influence of the electron-on-target energy and radial intensity FWHM on 40×40 cm^2 field profile doses from Tzedakis *et al.*[103] The calculated profiles are observed to be quite sensitive to these parameters. The central axis depth dose curves are also strongly influenced by the electron-on-target energy.[102] However, the central-axis depth-dose curves are quite insensitive to variation in the radial intensity distribution of the electron beam striking the target, because the dose along the central axis is deposited primarily by particles in the vicinity of the central axis. The divergence of the electron beam incident on the target also needs to be considered as it may affect large field profiles.

Regarding the influence of individual treatment head components, Sheikh–Bagheri and Rogers[102] (see Table II) showed that even small changes (0.01 cm) in the primary collimator's upstream opening can affect in-air off-axis ratios by restricting the number of bremsstrahlung photons contributing to the scattered photon fluence reaching off-axis points downstream. Dose profiles are quite sensitive to the composition and density of the flattening filter as noted in Fig. 4, where two different flattening filter materials were used for



(a)



(b)

FIG. 3. (a) The lateral dose-profile curves as a function of the FWHM of the radial intensity distribution for a monoenergetic, 6 MV, electron beam. The measured profile and five calculated dose profile curves are presented; each curve is normalized to the central axis. (b) The influence of electron energy on dose-profile curves. The energy was varied from monoenergetic 5.4 to 6.6 MeV. For clarity only four calculated and measured profiles are presented; each curve is normalized to the central axis. The FWHM of the radial intensity distribution was 0.3 cm FWHM in all cases. Reprinted from Tzedakis *et al.* (Ref. 103) with permission.

the calculations.[102,104] The careful specification of the geometry and materials within the MC code is an important consideration. This includes verification of the input of the component geometric specifications, for example, by using the listing files available with the BEAM (Ref. 23) system. Moreover, for complicated components, such as the flattening filter, independent attenuation calculations of these structures, using monoenergetic photons, may be helpful to verify their thickness profiles.[105,106] Attenuation of the photon beam by the monitor ion chamber and the field mirror is negligible and these structures are often omitted from MC simulations of photon beams, except when backscatter to the monitor chamber is being taken into account. The ion chamber and mirror are, however, potentially important structures in electron beam simulations. The collimating jaws have no significant effect on the energy and angular photon distributions.[14,107,108] Changing the composition of the secondary collimators from pure tungsten to an alloy containing 50% tungsten by weight does not significantly affect the calculated dose.[102]

Even in those cases for which the beam model includes an explicit simulation of the accelerator head, end users of MC treatment planning algorithms will probably have little control over parameters affecting the treatment head simulation. It will be incumbent upon vendors to provide accurate beam models (verified by measurements) for treatment planning purposes. Vendors and developers should be aware of potential difficulties associated with accurate specification of the treatment head component structures. Such issues include difficulty in obtaining proprietary information from accelerator manufacturers, incorrect proprietary information provided by the manufacturers,[101,102,104] undocumented accelerator updates, and large uncertainties in important parameters needed for accurate simulation, such as the electron-on-target energy. Accelerator vendors are encouraged to make accurate, detailed information of their accelerators accessible in formats easy to implement in MC simulation, as established for instance by Siemens.[109]

III.B.2. Electron beam specifics

Simulation of the passage of electron beams through the treatment heads of accelerators has figured prominently in MC radiotherapy calculations for many years. The first application of the BEAM code[23] was to simulate electron beams for a wide range of accelerators. A variety of publications have demonstrated the accuracy of this technique for computing dose distributions and output ratios.[24,110,111] Recent work has established the methodology to achieve high accuracy in matching calculated and measured dose distributions for even the largest fields, including asymmetries and the bremsstrahlung tail.[112]

The procedure for simulating the electron beam treatment head is quite similar to that of x rays. The components generally important for electron beam therapy treatment head simulations are shown in Table III. Photon beam components are also presented for comparison. For electrons, accelerators from the major vendors use a pair of scattering foils to flatten the beam with minimal bremsstrahlung contamination. A low-scatter monitor chamber may be employed and the mirror may be retracted from the beam. The jaw position is generally fixed for a given applicator and energy such that the relative output ratio depends only on the custom insert.

The critical parameters for electron beam simulation are different than those for x rays. Sensitivity analyses provide quantitative evidence of these differences.[100,101,113] Due to the sensitivity of the beam range to the primary electron energy (a 0.2 MeV change in electron energy corresponds with a 1 mm change in beam range), the incident electron energy is the primary tuning parameter for electron beam

TABLE II. A summary of the findings of Sheikh-Bagheri and Rogers (Ref. 102).

Parameter in the linac model	Impact on in-air off-axis factors	Impact on central-axis depth dose values
Mean energy of the incident electron intensity distribution (assumed Gaussian)	**Decrease with increasing primary electron energy**, e.g., -0.105 ± 0.007/MeV at 15 cm off-axis for a Siemens KD 6 MV beam. **A 0.2 MeV mean energy change produces an observable effect.**	**A 0.2 MeV change in mean energy causes an observable change** (2% or 3σ with 0.7% or 1σ dose uncertainty).
Gaussian width of the incident electron energy distribution	**Show little or no dependence**, e.g., widening of the FWHM from 0% to 20% resulted in no change, for a Siemens KD 6 MV beam. Asymmetrical energy distribution has a small effect, e.g., an asymmetric Gaussian with 14% FWHM on the LHS of the peak and 3% FWHM on the RHS of the peak causes a change of 2% for a Siemens KD 18 MV beam.	**Show weak dependence in the dose buildup region and at large depths**, e.g., an asymmetric Gaussian with 14% FWHM on the LHS of the peak and 3% FWHM on the RHS of the peak increases buildup dose by up to 1.5% for a Siemens KD 18 MV beam.
Gaussian width of the incident electron radial intensity distribution	**Decrease quadratically with increasing Gaussian width**, e.g., a change in FWHM from 0.01 to 0.15 cm leads to 7% decrease at 15 cm off-axis for a Varian 18 MV beam.	**Little or no observable effect** considering statistical uncertainties.
Divergence of the incident electron beam (at a given intensity distribution FWHM)	**Show little or no effect up to 0.5°**; at 1° show a decrease of 1% at 15 cm off axis at 100 cm SSD, for an 18 MV Varian beam.	**No observable effect up to a few degrees** considering statistical uncertainties (1% or 1σ).
Radius of the upstream opening of the primary collimator	**Sensitive to small changes**, e.g., varying the upstream opening by 0.01 cm produces a 1% change at 15 cm off-axis for a Varian 18 MV beam.	No observable effect.
Density and material of the flattening filter	**Show strong dependence**, e.g., reducing tungsten density by 1 g cm^{-3} causes a 6% reduction at 15 cm off axis for a Varian 15 MV beam. **Using the incorrect material has a very large effect**, primarily because of the density change.	Not reported.

simulations. However, electron distributions are very sensitive to all the materials in the beam, especially the scattering foils and may also be affected by the monitor chamber if it has thick walls. Hence, accurate geometric descriptions of all components in the beam path are required for the simulation of electron beams. Asymmetries are evident in electron dose distributions and parameters to adjust include angle of the incident beam and the lateral position of shaped scattering foils and the monitor chamber.

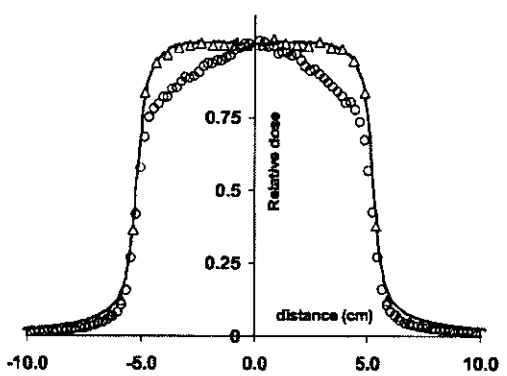


FIG. 4. MC calculated profiles in a water phantom (modified DPM, University of Michigan/UMPlan) for a 15 MV (10×10 cm^2) at 10 cm depth in water. Input for the profile calculations were the phase space simulations of the Varian 21-EX treatment head, performed using BEAMnrc, with two different material compositions specified for the flattening filter. Open circles represent the MC profile with a copper flattening filter (ρ =8.933 g/cm^3) and open triangles that with the tungsten filter (ρ =19.30 g/cm^3). Ion chamber measurements are shown in the solid line. Calculations and measurements are each normalized to the central axis (Ref. 104).

III.C. Modeling of the linear accelerator treatment head

III.C.1. General schemes

A beam model in the context of MC treatment planning is any algorithm that delivers the location, direction, and energy of particles to the patient dose-calculating algorithm. The direct MC simulation of a beam is one form of a beam model but for clarity we refer to it as a beam simulation rather than as a beam model. Accurate beam modeling is an important prerequisite for accurate dose calculation within the patient. Beam models use one of three possible ap-

TABLE III. Components in the treatment head and other beam modifying devices for x ray and electron beams.

X rays	Electrons
Exit window and target	Exit window and primary scattering foil
Primary collimator	Primary collimator
Flattening filter	Secondary scattering foil may be present
Monitor chamber	Monitor chamber (low scatter)
Mirror	Mirror
Asymmetric jaws and MLC	Jaw position fixed for each applicator
Wedge, blocks, graticule	Applicator with insert
Bolus	Bolus, shielding on or below patient surface

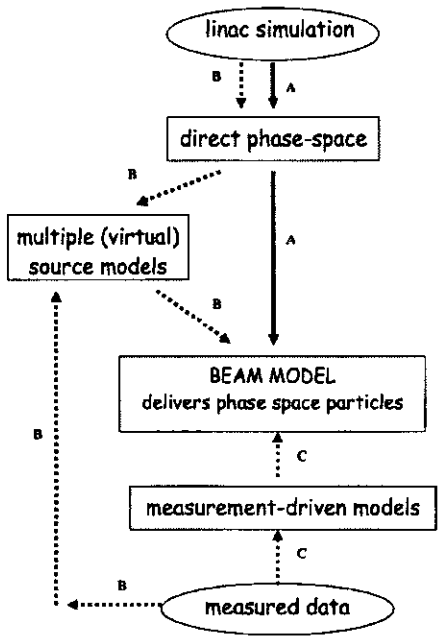


FIG. 5. The three "routes" for accelerator beam model specification: (a) solid line—direct use of phase-space information from simulation of the accelerator treatment head, (b) dashed line—multiple (virtual) source models derived from the phase-space information with or without enhancements from measured data, and (c) dotted line—development of other models derived from measurements (measurement-driven models).

proaches (see Fig. 5): (i) direct use of phase-space information from the accelerator treatment head simulation, (ii) development of virtual, multiple-source models reconstructed from the treatment head simulation with or without enhancement from measurements, or (iii) development of other models derived exclusively from measurements (measurement-driven models).

While the utilization of direct phase-space information provides details on the physical interactions within the treatment head, it may not be practical for routine clinical application.[65,93,114–116] Some of the limitations include: (a) generation and quality assurance of phase-space information requires MC simulation expertise, (b) accurate phase-space simulation is dependent upon accurate input parameters (such as the incident electron energy) as well as detailed geometric and material specifications of the accelerator components, which may be subject to inaccuracies, as well as proprietary issues, (c) storage requirements are large—typical MC simulation may require up to 10^9 phase-space particles for multiple photon beams,[115] which may amount to tens or hundreds of gigabytes of computer disk space for an accelerator with two photon energies and five electron energies with five different applicator sizes,[117] and (d) due to the much slower pace of speed increase in hard drives and networks compared to CPU speed development, reading the phase-space data can be a bottleneck in the calculation. It is therefore clear that a more concise, accurate characterization of radiation interactions in the treatment head is necessary for performing routine clinical MC dose calculation.

Another method for beam model specification, initially proposed by Ma *et al.*,[114] is based on the development of multiple-source models with parameters derived from the original simulated phase-space data.[65,88,93,114,116–120] In a multiple-source model, source particles are grouped by the location of their last interaction prior to being scored at the phase-space plane, resulting in subsources representing the major components of the treatment head. Fluence distributions for each subsource may be reconstructed from the phase-space data in the form of correlated histogram distributions,[65,93,114] thereby approximately retaining correlation of the particle's position, energy and direction. Although the development of accurate multiple-source models is also reliant upon simulation parameters and geometric and material specifications of the component structures, sub-sources may be "adjusted" (without redoing the simulations), based on measurements to optimize agreement between calculations and measured data, An example of such an approach consists of adjusting the spatial or energy boundaries of the planar fluence and energy distributions for each subsource to produce agreement with measured dose profiles and depth-dose curves, respectively, for a range of square field shapes, starting with fluence distributions reconstructed from phase-space data for one field size.[117,121] For instance, one changes the range of energies sampled from the discrete energy distribution as a function of radius, to provide a match with measurements, without having to redo the entire treatment head simulation. This method has been extended to MC-based commissioning of electron beams.[111,121] Models developed for reference accelerators were used to accurately commission (within 2%/2 mm of measurements in the high dose/high gradient profile regions) electron beams from other machines of similar design by adjusting a few parameters in the model. Implementation of photon and electron beam multiple-source models has been reported for a variety of accelerator manufacturers including, Varian,[122,123] Siemens,[73,124,125] Elekta,[73,86,88] Novalis (BrainLab),[87,89] and Cyberknife.[126]

A possibly more practical approach to beam modeling involves derivation of the model parameters from a standard set of measurements. The advantage of such measurement-driven models is that they may be developed without dependence on the details of the accelerator treatment head. Fluence distributions in measurement-driven models may be developed starting with analytical models whose parameters are optimized based on minimization of the differences between calculations and measurements.[100,125,127–131] Information for the model can be deconvolved from measured data.[100,127] Similar methods have been reported for beam-model specification using conventional dose calculation algorithms with good agreement established between calculations and measurements in water phantoms.[132–135] The same beam modeling criteria for conventional dose algorithms should also be applicable to the MC method although differences in the technical details such as source parameterization and software implementation are expected. The use of measurement-driven beam models eliminates the need for phase-space simulation and, as the model parameters are derived from measurements, may provide an added level of

confidence in the accuracy of the model in homogeneous phantoms. Moreover, specification of beam models is similar to that with conventional dose algorithms, and does not require expertise with MC accelerator simulation. However, rigorous verification of measurement-driven models in heterogeneous phantoms is necessary to validate the (empirically derived) fluence distributions. It has been shown that, in addition to the use of measurements in water phantoms, test cases in heterogeneous phantoms under conditions of electronic disequilibrium may be necessary to determine the correct energy spectrum unambiguously.[136]

III.C.2. Patient-specific beam modifiers

Transport through the patient-dependent components (such as the field-defining collimators and the MLC) may be classified into one of the following three schemes, which we will term: (a) explicit transport, (b) explicit-approximate transport, and (c) pseudo-explicit transport. In an explicit transport scheme, all particles (with appropriate energy cutoff values) are transported using MC techniques through the components; all details of the design geometry (such as rounded leaf ends, interleaf spacing, etc. for a MLC) should be included in the geometry modeling.[72,115,137] With explicit-approximate transport, approximations are employed in the MC photon/electron tracking scheme to improve the efficiency of the calculation.[83,87,138,139] An example is the approach of Siebers *et al.*,[83] in which only first Compton scattered photons are transported through the MLC. The method of Tyagi *et al.*[139] includes simulation through the detailed MLC geometry accounting for all Compton scattered photons, however, ignores the secondary electrons, i.e., assumes they deposit their energies locally. In a pseudo-explicit transport scheme, beam fluence distributions are reconstructed from the phase-space simulation to develop subsources for characterizing components, such as the field defining jaws,[93,120] electron applicators,[114] and the MLC.[88] Although the need for time-intensive, explicit transport is obviated with the use of multiple subsources, the ability to incorporate detailed geometric characteristics of components like the MLC may be difficult with this approach. Appropriate benchmarking must be performed by developers and vendors to evaluate the tradeoffs between speed and accuracy related to the use of various beam model approaches in MC-based modeling of patient-specific beam modifiers.

III.C.3. Output ratios

The output ratio (or relative output factor) is the ratio of dose per monitor unit in a phantom for an arbitrary field size to that for a reference collimator setting, the latter usually 10×10 cm^2 at a SAD (or SSD) of 100 cm for a depth of 10 cm for x-ray beams.[140] Output ratios over the range of clinically useful field sizes are heavily influenced by the amount of head scattered radiation impinging on the phantom.[140,141] Radiation generated by clinical accelerators can be characterized by a primary photon source generated through the bremsstrahlung process, and other extrafocal sources accounting for scattered photons arising primarily from the flattening filter and primary collimator, typically 3%–8% of the energy fluence in a 10×10 cm^2 field. The primary source is a narrow sharp source normally a few millimeters in diameter, while the extrafocal sources tend to be much broader, more diffuse and bell-shaped in nature.[142,143] The output ratio is affected by the size and geometry of these sources. The collimators (jaws and MLCs) block the extrafocal source when the field size is sufficiently small ($<3\times3$ cm^2) and for smaller fields the primary source starts being blocked by the collimators. The combination of extrafocal source and primary source cutoffs cause the output ratios to fall off sharply at small field sizes.[143,144] Accurate source modeling of small fields is especially important in IMRT planning where multiple, small field, off-axis segments are often used.[61,87,145] For medium sized fields (10×10 cm^2–20×20 cm^2), output ratios are more likely affected by the extrafocal sources as the collimators are large enough to expose the entire primary source, however, small enough to eclipse the extrafocal sources. In large fields, beyond 20×20 cm^2, the extrafocal sources are nearly completely exposed—the increase in output ratios in these situations is primarily a result of phantom scatter, lack of backscatter from the collimator jaws to the monitor chamber, and stray radiation from the treatment head.[146,147] Depending on the location of the jaws relative to the transmission chamber (which varies among the different linear accelerator types), radiation backscattered from the jaws into the chamber can also have an effect on the output ratios. Relative to large field sizes (40×40 cm^2), backscattered radiation increases by 2%–3% at small field sizes (3×3 cm^2) for 15 MV photons on a Varian (21EX) linear accelerator, for example. Studies for photon beam MC algorithms have shown calculated output ratios to be within 1.5% of measurements over a range of field sizes when accounting for backscattered radiation.[27,139,148,149] MC calculated electron beam output ratios (for field size-specific applicators) have also been reported to agree with measurements within 1%–2%.[26,110]

III.C.4. Dose buildup

There has been considerable discussion on discrepancies between MC calculations and measurements in the dose buildup region. Hartmann-Siantar *et al.*[27] observed MC dose deficits of up to 5 mm distance-to-agreement versus measurements and attributed these discrepancies to a source of electrons in the accelerator head not fully accounted for in the treatment head simulation with BEAM. It has been shown that arbitrarily increasing the electron contamination by a significant amount removes the discrepancy.[27,88,125] As the result of further investigation, Ding[150] concluded that the discrepancy could not be explained by the electron source hypothesis and postulated that contaminant neutrons emerging from the treatment head might be the cause. However, Ding *et al.*[151] subsequently refuted this neutron hypothesis by measuring minimal neutron component in the dose buildup region. Abdel-Rahman *et al.*,[152] who performed MC calculations (using EGSnrc) and a comprehensive set of measurements with multiple detectors, again found significant differ-

ences between measurements and calculations in the buildup region for 18 MV photons even when completely simulating the response of the detectors. They ruled out the following possible causes of the discrepancy: (a) an unknown electron source in the accelerator head simulation, (b) contaminating neutrons, (c) inaccurate cross section data, and (d) (gamma, p) reactions. They also showed that explicit modeling of triplet production interactions influences the dose buildup for an 18 MV beam. However, work by Kawrakow[153] showed that a more accurate triplet production model does not remove the discrepancies. Benchmarking of the NRCC accelerator photon beams has shown agreement well within 1%, for field sizes up to 10×10 cm^2, at all depths[154]; it should be noted that the NRCC accelerator (20 MV photons) uses a sweeping beam technique[154] as opposed to a flattening filter to flatten the beam. These comparisons explicitly accounted for stopping-power ratio variation with depth.

More recently, Kawrakow,[155] in performing detailed ion chamber simulations using the EGSnrc code system, showed that the relationship between measured ionization and dose for relative photon beam dosimetry depends on details of the chamber design, including cavity length, mass density of the wall material, size of the central electrode, and cavity radius, in addition to the beam quality and field size. When the correct ionization-to-dose relationship was used with the experimental data[155] and a variety of other improvements in the head simulations were made (e.g., using a larger diameter primary collimator opening and including the effects of extra shielding upstream of the jaws or MLC,[156] correcting a bug in the JAWS component of the BEAMnrc code, including an angular spread in the incident electron beam and several small effects in the simulation[153]), the discrepancies in the build-up region were reduced to an acceptable level. Chibani and Ma showed that resolving inaccuracies in the modeling of the primary collimator for a Varian 18 MV accelerator as well as including virtual sources for the lead shield and mirror frame, resulted in significantly better agreement between calculations and measurements in the dose buildup region.[156] Other sources of inaccuracy associated with detectors for dose buildup measurements are presented in Sec. III E 3.3.

III.D. Treatment planning: MC-based patient calculations

III.D.1. Statistical uncertainties

III.D.1.a. Latent variance and statistical estimators. For a finite number of independent simulated histories (N), the dose calculated using the MC method is subject to statistical uncertainty. By invoking the central limit theorem,[157] one can show that the statistical uncertainty in dose is proportional to $1/\sqrt{N}$, in the limit of infinite (large) N. There are generally two sources of statistical uncertainty in MC calculations of patient dose—those resulting from the simulation of the accelerator treatment head and those arising from fluctuations in the phantom/patient dose calculation. Sempau *et al.*[95] coined the term, "latent variance" to describe the uncertainty due to statistical fluctuations in the phase-space as opposed to the uncertainty due to the random nature of dose deposition in the phantom. Using an already calculated phase-space file, the statistical uncertainty in the dose calculated in a phantom by reusing the particles from the phase-space file (i.e., assuming they are independent and ignoring correlations between them), will approach the finite, latent variance associated with the phase space data, regardless of the number of times the phase space is reused. The use of source models derived from phase space simulation will tend to smooth out point fluctuations in the phase space.[114] However, if the latent variance is large enough to introduce systematic bias then this will be propagated in the reconstructed phase space (source model). Beam models derived exclusively from measurements, on the other hand, are analogous to those generated using conventional (analytical) algorithms—latent variance (as defined above) is not a concern for such models, but other systematic uncertainties in the beam model will be present. In estimating the statistical uncertainty in the patient dose calculation, it is necessary to account for the latent variance from the phase-space calculation as well as the random uncertainty from the patient calculation. To make this possible in practice, more work is needed to develop tools to assess the role of latent variance in patient dose calculations. Should latent variance be a significant factor in the total uncertainty, more independent phase-space particles need to be used in the patient simulations. It must be emphasized that all beam models are subject to systematic uncertainties, which are analogous to those introduced by the latent variance. For measurement-driven models, these uncertainties will be related to inaccuracies in the measurement data. Both types of models are subject to systematic uncertainties due to inadequacies in the model itself.

There are two common methods for calculating statistical uncertainties: the batch method and the history-by-history method. In the batch method, the estimate of uncertainty (standard error of the mean, $s_{\bar{x}}$) of a scored quantity, X, is given by

$$s_{\bar{x}} = \sqrt{\frac{\sum_{i=1}^{n}(X_i-\bar{X})^2}{n(n-1)}}, \tag{3a}$$

where n is the number of independent batches, X_i is the scored quantity (such as dose) in batch i, and $\bar{X}$ is the mean value of X over all the batches. The sample size is therefore given by the number of batches, where each batch is a calculation of the same quantity carried out with independent phase-space file inputs and random number sequences. In the history-by-history method, X_i represents the scored quantity in history i (rather than batch i) so that the standard error of the mean can be recast (in a mathematically equivalent form) as follows:

$$s_{\bar{x}} = \sqrt{\frac{1}{N-1}\left(\frac{\sum_{i=1}^{N}X_i^2}{N}-\left(\frac{\sum_{i=1}^{N}X_i}{N}\right)^2\right)}, \tag{3b}$$

where N is the number of primary (independent) histories, X_i the contribution to the scored quantity by independent his-

tory, i. An issue evident with the batch approach [Eq. (3a)] is that the sample size, n, is given by the number of batches. As n is usually small (on the order of ten or less) there is statistical fluctuation in the uncertainty itself. Advantages of the history-by-history method have been detailed elsewhere.[158]

An important consideration when calculating uncertainties is to take into account the correlation between a primary particle and all its secondaries, especially in the case of bremsstrahlung splitting where a large number of photons may all come from a single electron. Thus, to be strictly correct, these secondaries must be treated as part of the same history. If this correlation is not taken into account one can underestimate the uncertainty in a dose calculation, as the secondaries will be treated as independent particles thereby reducing the uncertainty erroneously. Another important case of correlation is when a single particle is being used several times as a source particle. In this situation it is important to recycle the particles, i.e., use them multiple times, all at once and treat them as part of the same history. If one were to restart the phase-space file multiple times, one would lose the correlation between particles which are all part of the same history. These secondaries would then be treated as independent particles causing the uncertainty to be underestimated. A more comprehensive review of the implications of recycling and restarting phase-space particles is provided in the paper by Walters *et al.*[158]

III.D.1.b. Influence of statistical uncertainties on dose distributions. For radiation therapy dose distributions, $s_{D_i} \propto \sqrt{D_i}$, where s_{D_i} is an estimate of the standard error of the mean (standard deviation/$\sqrt{N}$) of the dose in voxel i and D_i is the dose in that voxel.[159–161] The fractional (or relative) uncertainty in dose, $F_{D_i}=s_{D_i}/D_i \propto 1/\sqrt{D_i}$. In other words, the fractional uncertainty in dose in a voxel decreases as the dose increases. This relationship provides a useful rule of thumb when viewing dose distributions since it implies that the relative uncertainty in the dose in high-dose regions will be smaller than in low-dose regions, even though the absolute uncertainty is usually larger.

From Eq. (3b) we see that the uncertainty is roughly proportional to $1/\sqrt{N}$. Since, the simulation time $T \propto N$, it can be seen that achieving absolute precision ($s_D=0$) with MC simulation requires an infinite calculation time. Fortunately, absolute precision is not required in dose calculation results. This section concerns the precision required for MC simulation and the impact of a lack of precision on MC dose distribution quantities.

Radiation therapy dose distributions contain many voxels in which the dose is computed (for example, a cubic volume with sides of 10 cm contains 15 625 voxels with sides of 0.4 cm). Since the subvolume receiving a therapeutic dose can be highly variable, a standardized method to specify the statistical uncertainty for such a distribution is necessary. Although the statistical uncertainty can be specified for a single voxel in a dose distribution (such as that at the isocenter or at the maximum dose voxel, D_{max}), as described below, these voxels are poor measures for the uncertainty of a MC computed plan. Alternatively, the statistical uncertainty over some volume, such as a planning target volume or some dose volume, such as the volume receiving greater than $X\%$ of the treatment dose, can be computed from the square root of the average variance of each constituent voxel. For example, the fractional uncertainty in the average dose for voxels with dose values greater than 50% of the maximum dose, $\bar{F}_{D>0.5D_{max}}$ as suggested by Kawrakow, and Rogers and Mohan[99] could be used

$$\bar{F}_{D>0.5D_{max}} = \sqrt{\frac{1}{K_{D>0.5D_{max}}} \sum_{D>0.5D_{max}} \left(\frac{\overline{s_{D_i}}}{D_i}\right)^2}, \tag{4}$$

where D_i is the dose estimate in the ith voxel, $\overline{s_{D_i}}$ is its uncertainty, and the summation runs over the K voxels with dose greater than 50% of D_{max}. For clarity in communication, this report recommends that quantities, such as $\bar{F}_{D>0.5D_{max}}$ or $\bar{F}_{PTV}$ (or $\bar{F}_{PRV}$) for doses to the specific volumes, planning target volume (PTV) or planning risk volume (PRV), be adopted as a standard method of reporting statistical uncertainties in dose averaged over the relevant volume. The use of uncertainties to voxels, such as the maximum dose should be avoided. In situations where doses in single voxels are important, such as the maximum dose to a "serial" organ like the spinal cord, users are reminded to also consider the statistical uncertainty of that dose voxel.[162] In such instances, it may be necessary to simulate a large enough number of histories so that s_{D_i} is very small.[162] This will ensure that the absolute uncertainty in the highest dose voxel will also be small. The statistical precision required for dose estimates in single voxels should be decided upon with guidance from the clinician. Users should also note that for a constant number of source particles, the statistical uncertainty also depends upon the size of the dose voxel. Reducing the volume to achieve a "point-like" voxel will require increasing the number of particles simulated to achieve constant statistical precision.

To varying degrees, statistical uncertainties affect all measures of the dose distribution, including output ratios, isodose profiles, dose-volume histograms, dose response parameters such as equivalent uniform dose and TCP/NTCP. In a uniform dose distribution, the dose metrics most sensitive to statistical uncertainties are the maximum and minimum dose voxels. These extreme values are by definition the voxels that have the greatest deviation from the mean dose. If one desires a region of uniform dose (e.g., within the PTV), that dose will not be uniform using a MC-based calculation, due to statistical fluctuations between adjacent dose voxels. The reported minimum and maximum dose voxels will differ from the mean of the idealized dose distribution by up to several standard deviations if many voxels are present in the distribution.[45,163] For example, in a uniform dose distribution with 15 625 voxels computed with MC algorithms, due to statistical fluctuations, there is a 63% probability that the dose in at least one voxel differs from the mean by more than four standard deviations.[45] With regard to dose prescriptions, the specification of the maximum dose to a single voxel when there is a desired volume of uniform dose (e.g., the

PTV), will result in an underdosage to this volume. Similarly, dose prescription to the minimum dose voxel will result in an overdose to the relevant volume. Isodose contours are also sensitive to statistical noise. For well-defined fields, such as rectangular fields used in beam commissioning, even 1% (1s) statistical uncertainty causes observable jitter in isodose contours. For patient fields that have irregular shapes, the acceptable amount of statistical uncertainty for isodose viewing is a matter of personal preference and should be agreed upon by the planner and the physician. It has been suggested that 2% statistical uncertainty on the D_{max} voxel is adequate for isodose evaluation.[160] When viewing statistical jitter in MC isodose distributions, the physicist should remind observers that overall dose delivery accuracy is limited to within a few percent; therefore there is uncertainty in the actual location of an isodose surface even when dose is computed with non-MC algorithms. From this point of view, the planning team can use MC isodose jitter as a mechanism to open the dialog on realistic dose uncertainty in actual treatment delivery.

Integrated dose quantities, such as dose volume histograms (DVHs) are less sensitive to statistical uncertainty.[159–165] DVHs computed with the MC method represent the actual DVH (that computed with a hypothetical 0% statistical uncertainty simulation), convolved with a statistical uncertainty distribution. With this realization, Sempau and Bielajew[159] and Jiang *et al.*[165] reported that the effect of the "statistical noise" on DVHs could be removed by deconvolving the uncertainty from the DVH, allowing substantial decrease in the required number of histories used, depending on the complexity of the DVH. In general, the blurring effect due to statistical noise is greatest for steep DVHs, such as those for PTVs, while shallow DVHs, such as those for critical structures are less affected.

The sensitivity of quantities such as TCP and NTCP to statistical noise depends upon the parameters used by the model[164] and on the magnitude of the noise. Kawrakow[161] has shown for general dose-based cost functions that the uncertainty in the cost function decreases more rapidly than the individual dose uncertainties when the plan is close to optimum as expressed by the cost function. This implies that during IMRT optimization, where plan updates are based upon evaluation of such cost functions, larger statistical uncertainty might be acceptable.

A method to reduce the effect of statistical uncertainties in MC dose distributions is to postprocess the dose distribution. These methods have been termed denoising or smoothing techniques. Various methods related to digital filtering,[166,167] wavelet thresholding,[168,169] adaptive anisotropic diffusion,[170] and denoising based on a minimization problem[171] have been proposed. A detailed comparison of these methods is presented in the article by El Naqa *et al.*[172] Denoising is an approximate, efficiency enhancing method (i.e., it is not a variance reduction technique) since it can introduce systematic bias into the calculation. Nonetheless, denoising techniques are useful as they can reduce the overall (systematic +random) uncertainty when the random component decreases more than the systematic component increases. Kawrakow presented a series of tests to evaluate the suitability of MC dose distribution denoising algorithms.[167] These tests are based on the fact that one can determine the "correct" dose distribution by simulating many histories (for a very high precision) and then comparing the denoised distribution from a simulation with a much smaller number of histories to the high precision results. The tests include:

- visual inspection of isodose contours,
- evaluating root-mean-square difference between dose distributions,
- evaluating the maximum dose difference,
- comparisons of dose-volume histograms with and without denoising, and
- comparing the fraction of voxels failing an $x\%/y$ mm test.

Denoising methods reduce the number of particles (and, hence, the calculation time) required to achieve a given uncertainty by a factor of 3–10. Denoising techniques require proper validation under the full range of clinical circumstances before they are used with MC dose algorithms.

III.D.2. Dose prescriptions and monitor unit calculation

The stochastic nature of the MC method raises questions for prescribing dose. It is common clinical practice to prescribe dose to a single voxel or to base the dose prescription on the maximum or minimum dose voxels. However, as discussed above (Sec. III D 1.2), in an approximately uniform dose distribution, the outliers (the maximum and minimum dose voxels) are subject to the largest statistical fluctuation and even other single voxel doses may lack the precision for monitor unit calculations. Standard treatment planning analysis methods using isodose distributions and dose-volume histograms rely on dose averaged over a volume. It is logical to extend this practice to the prescription of dose, thereby avoiding precision issues in doses calculated in small volumes. For example, dose may be prescribed to an isodose surface, to a region of uniform dose (averaged over many voxels) about the isocenter, or to a single point on a dose volume histogram. The practice is emerging to calibrate the calculated dose distribution by performing a calculation in the standard geometry where the accelerator is set to deliver a given dose per monitor unit, (e.g., 1 cGy/MU) at a given voxel. In this case, the dose at the calibration point may be calculated to high precision by averaging over a large number of voxels in a uniform dose region.

The issue of monitor unit calculations to specific single voxels within the target volume (e.g., the isocenter) for routine clinical planning may be confounded by large statistical fluctuations in the doses to individual voxels. Although the treatment planning practice is quickly moving toward volume-based dose prescriptions, particularly for IMRT planning, the ability to perform second MU checks for plan verification purposes is an important component of the standard practice. Until more efficient solutions are available, MU cal-

culations based on single-voxel-dose prescriptions should be performed with very high precision. That said, this task group encourages a shift in paradigm from point-based toward volume-based dose prescriptions, which we feel will soon become the standard method for prescribing doses in radiotherapy planning. The task group strongly discourages vendors from using the D_{max} or D_{min} dose voxels or other single voxel doses for dose prescription and monitor unit calculations in their MC-based treatment planning systems. Current users of MC algorithms are encouraged to find ways of circumventing point-based dose prescriptions if their systems are not flexible enough to allow otherwise. To obviate the concerns of dose prescription based on a single voxel, one institution (author J.E.C., Ottawa Hospital) has developed the following "work around" method for MC-based electron beam calculations: A dose distribution is calculated using 100 MU; this dose distribution in absolute terms is equivalent to a relative isodose distribution normalized to the standard calibration conditions (10×10 applicator, 100 cm SSD, d_{max} depth). For dose prescription, the physician chooses the isodose line that encompasses the target. The dose prescription point is then positioned on this isodose line along the central axis of the beam, and the treatment MUs are determined. A second calculation is performed as a check of the MUs and the final dose distribution.

For multiple 3D-CRT or IMRT fields, the procedure for generating monitor units is similar to that for single fields. For example, a given isodose line (such as the 95% line) can be selected for dose prescription and, for a given field, the dose contribution to a single voxel (e.g., the projected field cax point) along this line is determined from the treatment plan. For a calibrated MC algorithm, the absolute dose contribution from the given field in this voxel (on the selected isodose line) will be computed in units of cGy/MU, from which the monitor units for the beam can be calculated. A complete formalism for MU calculations for the different types of treatment deliveries has been provided by Ma *et al.*[173]

III.D.3. CT-to-material conversions

For conventional algorithms, electron densities extracted from the CT image are used to scale the influence of primary and, ideally, also secondary radiation interactions. MC algorithms utilize material density and the material atomic composition when performing particle transport. The differing atomic compositions of patient materials (e.g., soft tissue, bone, lung, air) result in different cross sections for the various radiation interactions. While material compositions cannot be determined solely from a single energy CT, they can be indirectly approximated by estimating the mass density from the electron density followed by assigning a material to each voxel.[37,86,174–176] For some MC codes, explicit material specification is circumvented by directly relating the CT [Hounsfield (HU)] numbers to material interaction coefficients, based upon parameterization of materials representative of the patient,[37] for example those tabulated in ICRU Report No. 46.[177]

To ensure proper correspondence between HU and materials (or material interaction coefficients), correspondence between these quantities must be established during the CT and treatment planning system commissioning process. To ensure appropriate material specifications, it may be desirable to have several conversion tables, whose selection is based upon knowledge of the particular patient's characteristics. Use of multiple calibration curves reduces the volume of inappropriate tissues specified in given regions, such as lung tissue within a prostate gland, however. Although the importance of exact material specification has been established in other studies,[178] more work in this area of research relevant to clinical treatment planning is necessary. Other information about the patient may also be helpful. For example, in a patient with a hip implant, it may be impossible to distinguish if the implant is titanium, steel, or a composite based solely on the CT numbers. In this case, material assignment based on knowledge of the material implanted in the patient would be beneficial. A recent evaluation of the influence of material compositions on dose distributions[178] showed dose errors of up to 10% for 6 and 15 MV photons, and 30% for 18 MeV electrons due to media and/or mass density misassignment, when comparing dose distributions between a known phantom and a CT-imaged phantom with compositions and densities assigned by a conversion process. The use of conversion techniques based purely on mass density (e.g., assuming the only patient material is water, but with varying density), as employed in conventional algorithms, is discouraged with MC simulation because most of these methods ignore dependencies of particle interactions on the materials, which can lead to notable discrepancies in high atomic number materials.[176]

CT number artifacts caused by issues such as beam hardening in the CT scanning process or by high density structures, such as dental fillings are potentially important in MC dose calculation. Other artifacts may arise, for example, when the CT scanner encounters a sharp edge, such as the surface of a rectangular solid phantom, where a blurred edge may result after image reconstruction. In one observed case, the resultant contour showed 3 mm of additional phantom, causing the MC-calculated percentage depth dose curves to be shifted 3 mm toward the surface.[269] Such issues are of relevance to both MC- and non-MC-based algorithms and will need to be taken into consideration in order to perform accurate dose calculations in the dose buildup region.

As with any dose algorithm, testing should be performed to evaluate the effect of artifacts on the accuracy of the MC dose calculation.[179,180] It must be emphasized that the accuracy of CT-number to material conversions affects all dose calculation algorithms, both MC- and non-MC-based methods.

III.D.4. Dose-to-water and dose-to-medium

Historically, radiotherapy dose measurements and calculations have been performed in, or specified in terms of the absorbed dose to water (D_w). With MC-based algorithms,

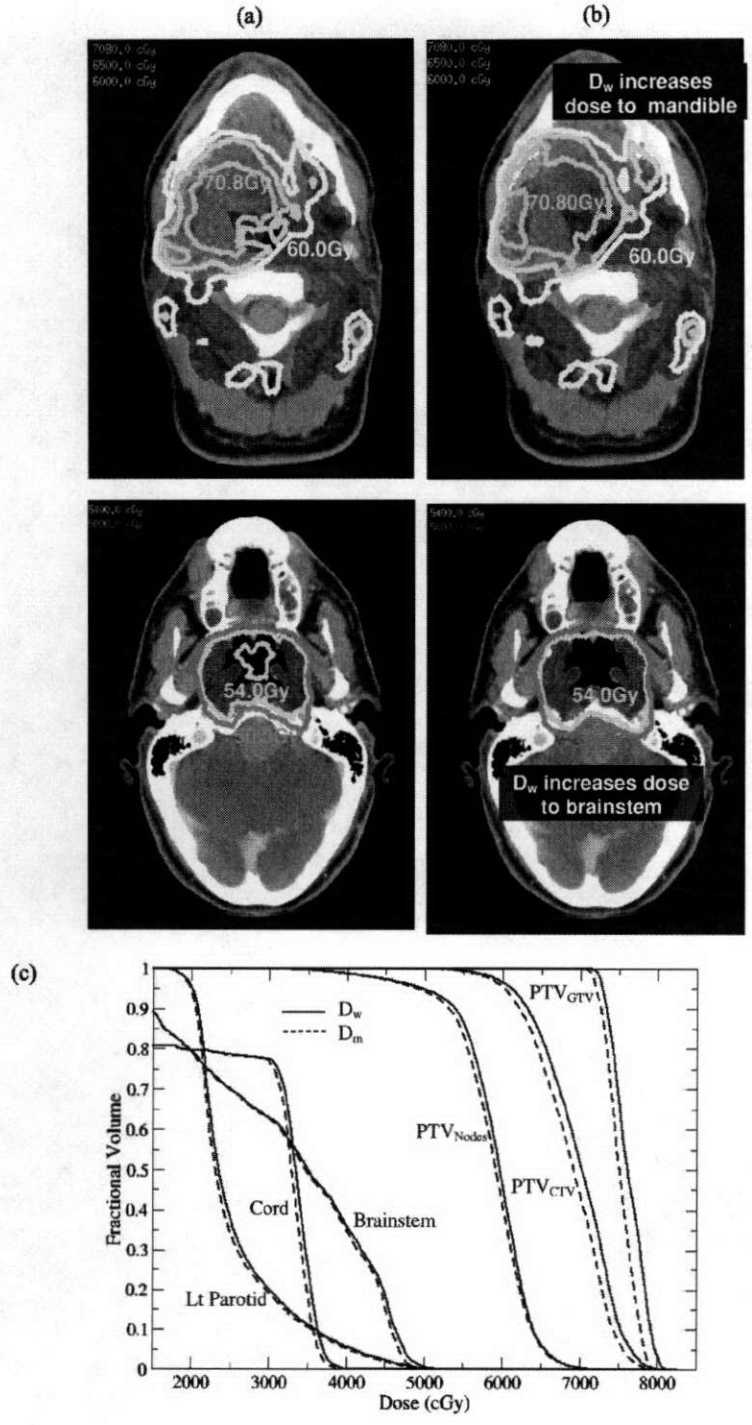


FIG. 6. MC-based (MCV system) dose-to-medium (D_m) and dose-to-water (D_w) results for a typical head and neck IMRT treatment plan: (a) D_m, (b) D_w, and (c) DVH comparison. Reprinted from Dogan *et al.* (Ref. 181) with permission.

particle transport simulations occur in materials representative of patient media; dose is therefore specified to the patient medium (D_m). For tissues with densities near 1.0 g/cm^3, the difference between D_w and D_m for megavoltage photon beams is small (1%–2%), however, for higher density materials, such as cortical bone, the difference can be as large as 15%,[176] since the stopping powers of water and these higher-density materials differ more significantly. Therefore, there is a systematic difference between the dose computed using conventional analytical algorithms and MC simulation. Figure 6 shows the dose and DVH differences between plans calculated with D_w and D_m for a typical head and neck IMRT treatment plan.[181]

To use MC simulation in the current clinical practice so as to be able to compare D_m with historical D_w results, requires a conversion of D_m to D_w for dose prescriptions, isodose coverage, dose-volume histograms, and any other dose related metrics. In this context, the converted D_w represents the dose to a small volume of water embedded in the actual medium. Whether one should eventually use D_m in place of

D_w directly in clinical prescriptions remains the subject of debate.[182] Arguments in favor of using D_w for dose specifications include:

- Historical clinical experience has been derived based on D_w, hence, D_w allows direct compliance with previous clinical experience and with conventional dose algorithms. Doses reported in clinical trials are based on D_w, hence, therapeutic and normal tissue tolerance doses are based on D_w.
- Accelerator and ionization chamber calibration protocols are D_w based.
- Tumor cells embedded within a medium are more water-like than medium-like, e.g., a tumor cell embedded in a bone matrix.

Arguments in favor of using D_m for dose specification include:

- D_m (or the dose to the tissues of interest) is the quantity inherently computed by MC dose algorithms. This may be of more clinical relevance than the doses on which historical clinical experience is based, which are approximate estimates of the true dose in the first place.
- Converting D_m back to D_w may involve additional complexity and introduce additional dose uncertainty.
- The difference between D_m and D_w for tissue equivalent materials is rather small and is likely to have minimal impact in clinical practice.
- It is known that, due to organ and target motion, the dose actually delivered in the course of a treatment may be significantly different from the planned dose. Conversion of Monte Carlo computed doses to D_w may analogously necessitate discussions on conversion to "dose to a static patient," should 4D planning and delivery techniques become routinely used in the clinic.[183]

The conversion of D_m to D_w or vice versa requires an application of Bragg–Gray cavity theory: $D_w = D_m(\bar{S}/\rho)^w_m$, where $(\bar{S}/\rho)^w_m$ is the unrestricted water-to-medium mass collision stopping power averaged over the energy spectra of primary electrons at the point of interest. One method to accomplish this conversion[176] is to utilize the fact that for patient-like materials, $(\bar{S}/\rho)^w_m$ is approximately invariant throughout a photon radiation therapy field (within 1%), hence, $(\bar{S}/\rho)^w_m$ can be used as a post-processing step to convert D_m to D_w. However, this requires that a sufficient number of materials be specified so that the material-dependent dose conversion approximates a continuous function.[184] Rather than postprocessing the dose conversion, one can multiply the energy deposited by primary and secondary electrons on each electron energy-loss step by the factor $(L/\rho)^w_m$, the ratio of the restricted mass collision stopping powers of water to local medium for the current energy of the electron. This is done in the MC transport code, thereby directly obtaining D_w. Alternatively, material interaction data may be evaluated over continuously parameterized space with cross sections and stopping powers for specific materials scaled relative to water.[35,37] Figure 7 shows a parameterization of the mass collision stopping power as a function of tissue density for different electron beam energies.[37] The fit is shown to be in good agreement with ICRU data[177] for body tissues. The advantage of this approach is that it circumvents the potential misassignment of media at material boundaries[37] which arises when the human tissue mass density range is divided into bins representing different tissues, especially when the number of material bins is small.[184]

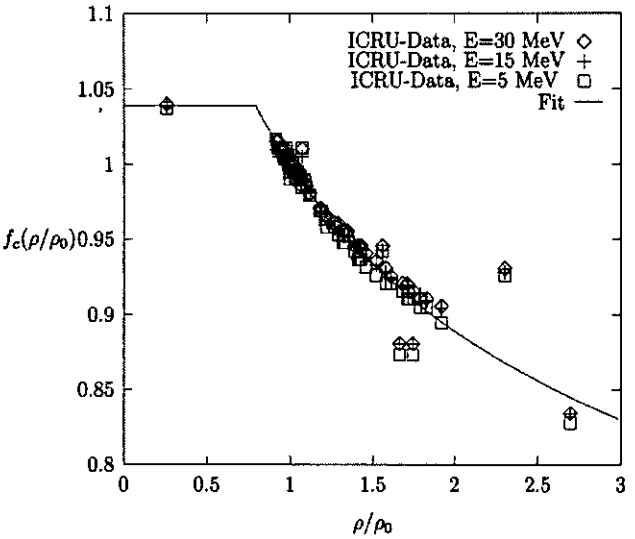


FIG. 7. Mass collision stopping power divided by the mass collision stopping power of water as a function of density normalized to water. A comparison of the fit function to the ICRU data (Ref. 177) for body tissues for electron energies of 5, 15, and 30 MeV is shown. The few points outside the curve are materials like urinary stones, which are considered to have negligible effects. Reprinted from Kawrakow *et al.* (Ref. 37) with permission.

Until further studies indicate clinical justification for selecting D_m or D_w, it is the consensus of this task group that MC dose results should: (a) explicitly indicate the material to which the dose is computed and (b) allow conversion between D_m and D_w using one of the methods discussed above, or other methods as developed in future investigations.

III.D.5. IMRT dose calculation and optimization

A strength of the MC method is its ability to accurately compute dose for complex dose delivery scenarios such as encountered in IMRT. IMRT often involves large intensity gradients and is usually delivered using a sequence of small static or dynamically shaped MLC segments. Under these circumstances, the assumptions used in conventional algorithms regarding scatter equilibrium and output ratio variation with field size often break down.[185] Additionally, in IMRT a significant fraction of dose to structures of interest (particularly dose limiting critical structures) is due to radiation scattered from or transmitted through the MLC.[186] MC simulation circumvents these limitations since it makes no assumptions regarding radiation equilibrium and can transport particles through the detailed MLC leaf geometry[87,137,187] to include MLC leakage radiation, even for dynamic MLC leaf sequences.[138,188,189]

In using MC calculation in IMRT, a consideration is the method used to incorporate the MLC into the dose calculation. Intensity modulation has been incorporated into MC simulation using the conventional planning system's intensity matrix,[190] independently generated fluence modification matrices,[87,92,191,192] or direct transport through the MLC.[61,189] Note that errors introduced by fluence approximations used during the MC dose calculation will be propagated through to the prediction of the patient's dose. Thus, when a fluence matrix approach is used for MC dose calculations, differences with respect to a conventional algorithm's heterogeneity correction will be detected, but fluence prediction errors may go undetected, particularly if the same fluence matrix is used both for the conventional and MC calculations. However, when MC simulation is used to transport directly through the detailed MLC geometry, these fluence errors should be detected. Accounting for geometric details in the MLC geometry, such as interleaf leakage is possible using a fluence matrix approach, however, will require a very high resolution calculation matrix, which may be a limiting factor. Moreover, the energies of the scattered particles through the MLC using such an approach will be approximate. Whether or not modeling the intricate details of the MLC (versus more approximate fluence matrix approaches) will lead to clinically significant differences in IMRT treatment planning is not a fully resolved issue. More treatment planning studies in this area of research are necessary to better understand the associated clinical implications.

Application of the MC method for IMRT QA has been demonstrated by several research groups[61,87,139,190,192–194] using MC simulation to recompute dose distributions optimized with the conventional dose planning algorithm. When using MC calculations as the reference plan and ignoring statistical fluctuations, dose differences between conventional and MC algorithms can be considered systematic.[195] Ma *et al.*[191] found dose errors in excess of 5% and 20% (relative to the prescribed dose) in targets and critical structures respectively due to patient heterogeneities, in comparing MC calculations (employing an independent fluence matrix) with a pencil-beam model. In comparing MC calculations with a conventional planning system's (pencil-beam model) intensity matrix for head and neck and lung cases, Wang *et al.*[190] found a 20% lower V_{95} (volume receiving at least 95% of the target dose) for a lung plan, and a 9% lower D_{95} (dose delivered to at least 95% of the target volume) for a head and neck plan. Average agreement among all head/neck and lung cases in this study, however, was quite good.[190] Regarding the transport of particles through the MLC, Siebers and Mohan[61] showed fluence-based IMRT dose underestimates of 4.5% in V_{95} and heterogeneity-based dose overestimates of 5% in V_{15} in the same treatment plan. In intercomparing MC codes, Reynaert *et al.*[196] reported deviations of up to 10% in DVHs between PEREGRINE and BEAMnrc calculations. These discrepancies were attributed to differences in modeling of the Elekta (SLi-plus) MLC, not caused by the particle transport in the patient. This study further illustrates the importance of careful modeling of the details of the MLC in IMRT treatment planning.

MC-based IMRT plan optimization is limited by the clinical availability of MC calculation as a whole and the large calculation time required to perform the multiple IMRT dose calculations required for optimization. MC simulation during optimization allows the optimizer to account for heterogeneity induced dose perturbations, as well as for MLC leakage and scattered radiation. Inaccurate dose algorithms used during optimization can result in convergence errors[195] in which the optimized fluence pattern differs from that corresponding to the optimal dose distribution.

Studies demonstrating the use of the MC method in IMRT optimization include the work of Laub *et al.*,[197] who utilized a MC algorithm to evaluate the cost function during optimization but a pencil beam to compute the cost function derivatives used by the optimizer. Jeraj *et al.*,[195] reported on convergence and systematic errors in the IMRT inverse planning process (for lung cancer) resulting from the use of convolution/superposition and pencil beam algorithms, versus the Monte Carlo method. A similar study for treatment planning of head and neck cancers was published by Dogan *et al.*[198] Another study by Siebers *et al.*[199] demonstrated the use of correction-based schemes to produce convergence of doses computed by conventional algorithms with MC calculations. Bergman *et al.*[200] reported on the use of an EGSnrc-based MC beamlet dose distribution matrix for IMRT planning using a direct aperture optimization algorithm. The goal of their work was to assess the improvement in accuracy over conventional algorithms in using MC methods for both the final dose calculation as well as in the inverse planning process.[200] Combining these methods with statistical smoothing and denoising techniques,[166–172] after comprehensive benchmarking, may allow introduction of MC-based IMRT optimization into routine clinical practice, particularly since it has been shown that cost functions converge faster than individual dose uncertainties.[161] For a review of other studies, the reader is referred to the article by Verhaegen and Seuntjens.[25]

III.D.6. Voxel size effects

As is the case with any dose calculation algorithm, calculated dose is affected by the size of the scoring voxel. For MC calculations, typical values in the scoring dimension are voxel sides of 2–5 mm for field sizes greater than 3×3 cm^2 and 1–2 mm for field sizes less than 3×3 cm^2. For calculations where geometric details of the MLC are included in the modeling, scoring voxel sizes no larger than 1–2 mm will be necessary to diminish volume averaging of dose from inter- and intraleaf leakage. As with conventional algorithms, MC-based IMRT calculations should be performed using voxel sizes of 2–3 mm or less in the high gradient regions.[201,202] In addition to affecting the spatial resolution, the statistical uncertainty will be influenced by the voxel size; reducing the voxel size will increase the relative uncertainty for a fixed number of source particles because fewer particles deposit dose in the smaller volume. Increasing the voxel size (and, hence, volume) will reduce the relative uncertainty but may introduce errors due to re-

duced spatial resolution. An example of the influence of volume-averaging effects resulting from the use of larger voxel sizes (~0.5 cm on each side) in MC electron calculations is discussed in Sec. III E 3.5.

III.D.7. Cross sections

The uncertainties in photon interaction cross section data in the energy range from 5 keV to a few MeV is of the order of 1%–2%.[203] Although many MC codes use the incoherent-scattering-factor approximation (which assumes scattering of photons from stationary, free electrons), this approximation is found to be accurate in the mega-voltage energy regime, where the energy of the incident photon is much higher than that of the electron *K*-shell binding energy.[204] An excellent review of the cross sections for bremsstrahlung production and electron-impact ionization has been provided by Seltzer.[205] It is shown that, in general, calculated cross sections for the various electron interaction processes are in agreement with measurements within the combined experimental and theoretical uncertainties.[205] It is felt that cross section and electron data in the megavoltage energy range are sufficiently accurate,[206] assuming that sampling of these cross sections is done accurately within a given code. These same effects will be present in doses computed with the convolution/superposition algorithm.

III.E. Experimental verification

III.E.1. Introduction

In this report experimental verification of the MC algorithm deals with how accurately the algorithm performs under different test conditions within a phantom. As with any algorithm, verification and testing is a necessary step to ensure safety of use in the clinical setting. It is the consensus of this task group that verification of a MC algorithm should be similar to that of any model-based dose calculation algorithm, such as convolution/superposition. The clinical commissioning and acceptance testing of dose calculation algorithms has been reported.[48–50] Additional testing to confirm the accuracy of the MC algorithm in situations of electronic disequilibrium will be helpful. Quantification of benchmark cases (such as the ICCR benchmark[99]) should be performed by either the user, or the vendor. The intent of this section is to provide some examples of additional types of testing that may be included to assure the accuracy of the MC algorithm. The specification of required measurements for acceptance testing and commissioning of the MC algorithm and the criteria for algorithmic agreement with measurements is beyond the scope of this report.

III.E.2. Previous work

Toward the goal of verification of dose calculation algorithms, there have been a variety of studies related to measurements in heterogeneous media. These investigations (see, for example, Refs. 207–212) have usually focused on establishing limitations of conventional dose algorithms in heterogeneous media and on how the use of physics-based algorithms, such as the convolution and MC methods, can be used to produce more accurate results. More recently, Arnfield *et al.*[213] found large differences (up to 10%) between collapsed cone convolution and MC (PEREGRINE) calculations in an 8 cm lung-equivalent slab embedded within solid water and irradiated by a 4×4 cm^2, 18 MV photon beam. Their work included film and ion chamber detectors and showed that MC calculations were in good agreement with measurements.[213]

With respect to patient planning, studies[91,115,214–219] have pointed out major differences between MC calculation and conventional methods, such as the 3D pencil beam and convolution/superposition algorithms. Over the past ten years, there has been a growing interest in the use of the MC method in clinical treatment applications with many institutions around the world actively involved in the development and testing of such systems. Many experiments in homogeneous and heterogeneous phantoms[26,27,71,75,82,84,86,93,96,115,119,194,213,220–225] have been directed toward verification of MC algorithms for clinical planning applications. The interested reader is referred to these and other related publications for a comprehensive review of the various types of experimental testing of MC treatment planning algorithms.

III.E.3. Types of verification experiments

Experimental verification of a MC algorithm should include testing to assess the accuracy of: (a) the beam model (be it measurement-driven or based on treatment head simulation) and (b) the radiation transport algorithm in homogeneous and heterogeneous phantoms. The former is part of routine commissioning of dose calculation algorithms, whereas the latter is likely to have significantly more involvement from developers and vendors.

III.E.3.a. The beam model. The purpose of verification of the beam (treatment head) model is to ensure that parameters, such as the incident beam energy (if used) are correctly "tuned" to produce dose distributions in agreement with measurement. Such verification is the same as that for any dose algorithm and may best be performed with measurements in a homogeneous (water) phantom. These tests should include the acquisition of depth and profile doses in a water phantom for a range of field sizes, as is routinely performed for conventional algorithmic verification, and documented in the AAPM TG-53 report.[48–50] In addition, the use of measured in-air off axis ratios may be useful for benchmarking the beam model. Calculated in-air off-axis ratios have been shown to be very sensitive to the incident electron beam parameters (e.g., mean energy, intensity distribution, etc.), as well as the dimensions and densities of other structures, such as the primary collimator and flattening filter.[102]

The multileaf collimator (MLC). Experiments to benchmark the MLC transport have ranged from arbitrarily shaped AP fields designed to verify overall penumbral and transmission dose[88] to more complicated MLC shapes designed to test modeling of detailed effects, such as transport through the rounded leaf ends,[83,87,226] tongue-and-groove

effect,[27,72,83,87,137] and intra/interleaf transmission.[27,72,83,87,137] The complexity of MLC verification experiments will depend on the usage of the MC algorithm in the clinical setting. For the purposes of 3D conformal radiotherapy (3D-CRT), the shaped field cases provided in Fig. A3–1 of the AAPM TG-53 (Ref. 48) report are good examples of tests designed to test overall MLC transmission and penumbral effects. In the context of 3D-CRT planning, the accurate modeling of details of the MLC and the influence of these effects on leaf transmission and penumbra may be of reduced clinical importance to the target doses (for an appropriate CTV-to-PTV margin) as these issues affect the dose at the field edges and outside the field.[227] For IMRT, it is now well established that the accurate modeling of MLC transmission and penumbra is critical.[61,227] A stringent test of the ability of the MLC model to accurately handle intra/interleaf transmission is presented in Fig. 8(a), for a Varian, Millenium 120-leaf MLC. In this example, the calculated MLC leakage radiation is compared with film measurements in a direction perpendicular to the leaf motion.[83] Figure 8(b) shows a comparison between MC calculations and film measurements for a MLC "picket fence" shape where the field is blocked by even numbered MLC leaves with odd numbered leaves retracted behind the jaws.[83] This test is useful in evaluating how accurately the tongue-and-groove effect is handled.

In designing a test suite for verification of the MLC transport accuracy, the clinical physicist should give special consideration to the detailed specifications of the MLC, particularly if the MC model is to be used for IMRT planning. These details include: density and composition of the MLC leaves, rounded leaf-end dimensions, intra/interleaf gaps, and tongue-and-groove dimensions. Discrepancies between MC calculations and measurements deemed significant by the physicist should be reported to the vendor; modification of the relevant parameters influencing the model accuracy, such as the MLC geometric model should be performed by the vendor to establish acceptable agreement between calculations and measurements.

Other beam modifying devices. The accuracy of MC simulation of beam modifying devices, such as wedges and blocks may be benchmarked using methods similar to that for conventional algorithms.[48] A publication on the experimental verification of the PEREGRINE MC system[27] included the acquisition of large wedged field profiles (40×20 cm^2) at various depths (see Fig. 13 of Hartmann-Siantar *et al.*[27]). The purpose of large field testing is to ensure that radiation transport through the wedge (or other beam modifying device) is in acceptable agreement with measurements across the entire physical range of the device. Unacceptable differences between calculations and measurements should be reported to the vendor—appropriate action should be taken by the vendor to help resolve these differences.

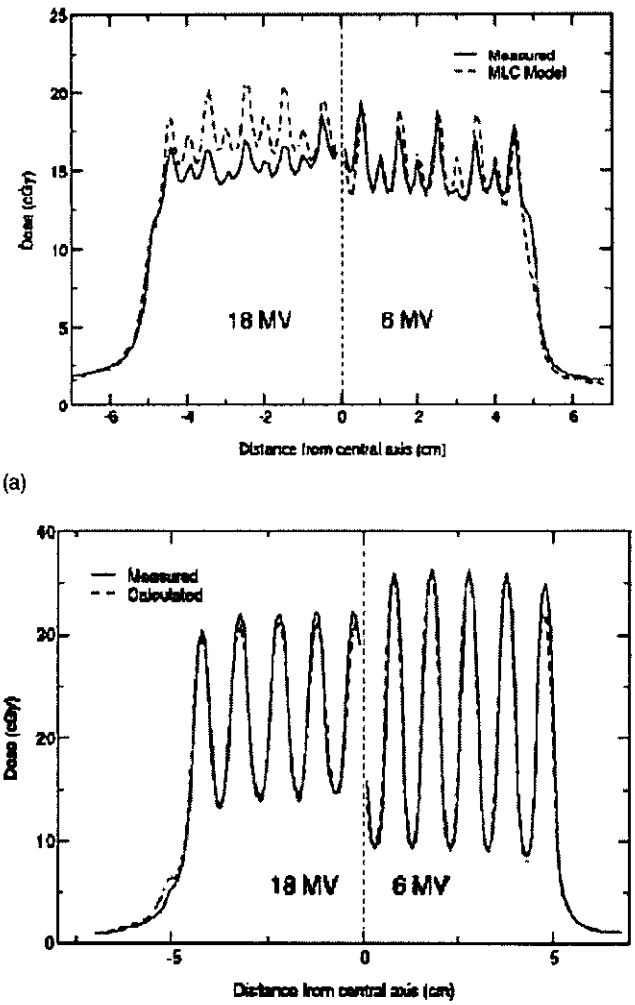


FIG. 8. (a) Measured (solid line) and calculated (dashed line) MLC leakage radiation perpendicular to the direction of MLC leaf motion for a 10 $\times$ 10 cm^2 MLC-blocked field for 6 and 18 MV beams. Dose calculation and measurement for the 6 MV beams occurred at 5 cm depth, 95 cm SSD, and the 18 MV data at 10 cm depth, 90 cm SSD. One should note that the dose due to leakage radiation (under the closed leafs) typically accounts for 2% – 3% of the open field dose. The discrepancy noted for the 18 MV comparison is roughly 0.1% of the open field dose and is most likely to a small difference in the MLC density used in the calculations (Ref. 83). Reprinted from Siebers *et al.* (Ref. 83) with permission. (b) Measured (solid line) and calculated (dashed line) doses for 6 and 18 MV 10×10 cm^2 field (a) blocked by even numbered MLC leaves with odd numbered MLC leaves retracted behind the jaws. One should note that the dose due to leakage radiation (under the closed leafs) typically accounts for 2% – 3% of the open field dose. Reprinted from Siebers *et al.* (Ref. 83) with permission.

III.E.4. Verification of the Monte Carlo transport algorithm in phantom

Given the improvement in the accuracy of MC simulation over conventional algorithms, particularly under circumstances of electronic disequilibrium, verification of the MC algorithm should include testing under these types of conditions to confirm the expected accuracy. Although reports, such as the AAPM TG-53 (Ref. 48) and others,[49,50] recommend testing in heterogeneous phantoms, issues related to electronic disequilibrium are generally excluded. This task group strongly encourages that verification testing of the MC algorithm include experiments emphasizing electronic disequilibrium effects, in addition to standard tests in heteroge-

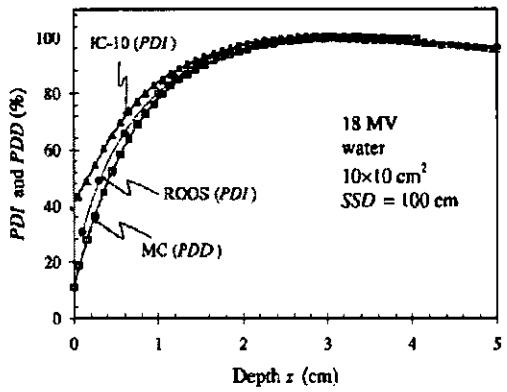


FIG. 9. Measured percent depth ionization and Monte Carlo calculated percent depth dose in the dose buildup region in water with the IC-10 cylindrical ionization chamber and the Roos parallel-plate ionization chamber in water for a 10×10 cm^2 field and SSD of 100 cm, for an 18 MV x-ray beam. For the IC-10 ion chamber, the effective point of measurement was 1.8 mm upstream from the chamber center. Reprinted from Abdel-Rahman *et al.* (Ref. 152) with permission.

neous phantoms as documented in other reports.[48–50] This type of testing is needed to exploit the advantages of MC-based simulation to its full potential.

Verification measurements in slab phantoms with embedded low density inserts and irradiated by high energy photons are useful in assessing the transport accuracy under conditions of electronic disequilibrium.[84,213,220] Effects of penumbral broadening of the dose distribution resulting from lateral electron transport may be evaluated using such phantoms.[84,213,220]

Dose measurements at various locations (depths and off-axis distances) both within and outside a high-density heterogeneity provide a benchmark to evaluate the perturbation of the photon and electron fluence due to the presence of the heterogeneity.[225] Dosimetric effects at interfaces (e.g., due to backscattering at bone/tissue interfaces[53,225]), transport within the heterogeneity, and the doses outside the heterogeneity may all be assessed with careful positioning of measurement devices within the heterogeneous phantom.

Experimental verification should also be performed in more clinically relevant situations. Examples of such geometries include thorax phantoms,[228] and mediastinal[229] and tumor-like phantoms.[229,230]

Verification experiments for algorithmic verification have also included the use of anthropomorphic (Rando) phantoms, where TLDs are most often used for the measurements.[225,231] With anthropomorphic phantoms, calculated patient treatment plans may be verified under a range of clinical circumstances. However, measurements must be carried out carefully with appropriate experimental techniques, theoretical interpretation, and reproducibility.

III.E.5. Dose buildup region

It is challenging to make accurate measurements in the dose buildup region. A recent study by Abdel-Rahman *et al.*[152] showed significant differences (see Fig. 9) in the dose buildup region (<1 cm) between the response of a parallel plate (Roos) and cylindrical ion chambers (IC-10, Scanditronix) in both 6 and 18 MV photon beams. Similar differences were found between a cylindrical ion chamber and a parallel plate chamber (P11) and a stereotactic diode (Scanditronix) by Yokoyama *et al.*[232] who measured dose buildup regions for IMRT fields. Based on the available literature, it is therefore recommended that dose buildups be carefully measured with either a parallel plate chamber, an extrapolation chamber, or with methods such as TLD extrapolation.[233] Diode detectors may also be considered for dose buildup measurements[234] but should be cross referenced with other detectors, such as parallel plate ion chambers. As pointed out by Kawrakow[155] (see Sec. III C 4), the relationship between the measured ionization and dose is sensitive to details of the ion chamber design and needs to be accounted for in dose buildup measurements.

III.E.6. Output ratios

MC calculations of output should be performed in a measurement-like geometry, usually consisting of a central axis point dose estimate at a fixed depth (d_{max}, 10 cm or other) in a water phantom for square field sizes defined by the collimating jaws. If the backscattered radiation into the monitor chamber is correctly modeled, it is possible to calculate output ratios to within 1%–2% agreement with measurements over a range of square field sizes, with sides from 3 to 40 cm.[146] In addition, comprehensive verification of output ratios should include testing for small segmental fields located off-axis, which are commonly used in IMRT planning. Output ratios should be verified against measurements of specially designed IMRT fields, such as junction narrow slit fields, in which the effects of small field dosimetry are magnified. Examples of such fields have been reported.[83,87,145,187]

III.E.7. Electron beams

Careful experimental verification is especially necessary for electrons because the tolerance of parameters, such as the electron energy and the treatment head structure constituents are much tighter than for photons. Accurate measurements are a prerequisite to accurate simulation. Care must be taken in measurement of the central axis depth-dose curve used to define the beam energy. Profiles in large fields need to be included to determine geometry details such as the distance between the scattering foils, the thickness of these foils, and the lateral position of shaped scattering foils, if present. Measurement of dose profiles in the bremsstrahlung tail are helpful to validate the photon component of the beam model.[112] Comparison to measured output ratios and dose distributions over the full clinical range of field sizes and SSDs is necessary for each applicator and beam energy, to rigorously validate the dose calculation. Further measurements, such as dose measured in an anthropomorphic phantom,[221] should be performed to validate the beam model for the specific MC-based application.

Comprehensive verification for electron beam MC simulation should include measurements in heterogeneous phan-

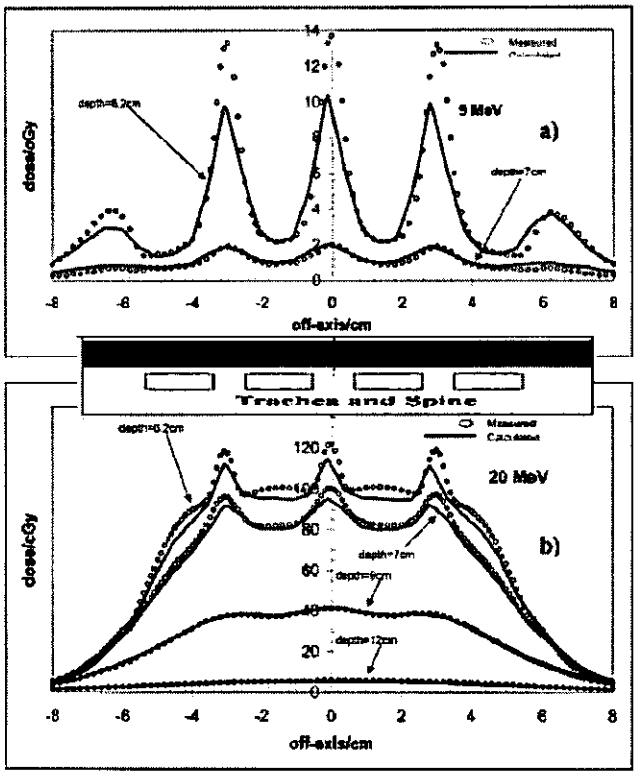


FIG. 10. Trachea and spine phantom: measured and Monte Carlo (Masterplan, Nucletron, based on VMC++) calculated crossplane dose profiles at various depths for: (a) 9 and (b) 20 MeV. SSD=100 cm, 10×10 cm^2 applicator. Monte Carlo simulations were performed with 50 000 histories/cm^2 and a voxel size of 0.49 cm. 100 MU were used for the calculations. The relative uncertainty in the calculations is about 1% −1.5%. The phantom geometry is shown in the inset. Differences between measurements and calculations in this example were attributed to the volume averaging effects of a large voxel size (~0.5 cm) (Ref. 26). Reprinted from Cygler *et al.* (Ref. 26) with permission.

toms, such as those reported in the Electron Collaborative Working Group report[235] and its update.[236] Cygler *et al.*[26] performed experimental benchmarks using a variety of heterogeneous phantoms, including a 1D slab geometry with an aluminum slab insert, a 2D geometry with cortical bone-equivalent inserts ("ribs" geometry), a 3D cylindrical geometry with an "air" insert, and a complex trachea and spine equivalent geometry. Figure 10 shows the example trachea and spine phantom used in their study.[26] Differences between measurements and calculations in this example were shown to be due to the volume averaging effects of a large voxel size (~0.5 cm).[26] The MC calculations are found to underestimate the measurements because the large voxel size averages the density distribution in the vicinity of the heterogeneity, effectively reducing the calculated dose across the low-density heterogeneity.[26]

Careful measurements in such heterogeneous phantoms will provide rigorous benchmark data for verification of the MC algorithm, and as such should be strongly considered for testing purposes.

III.E.8. Measurement uncertainties

Making dose measurements is fraught with problems if the aim is to verify calculations at the 2% or better level, particularly in nonuniform dose regions. It is important to take into account point of measurement effects, variations in stopping power ratios for ion chambers, and beam quality dependence in general for other detectors (e.g., the well known over-response of film in regions of low mean photon energy), polarity effects, ion recombination effects, and chamber perturbation effects. Specific detectors may be required for accurate measurements under specific conditions; for example, properly constructed parallel plate chambers may be favored over ion chambers for measurements in the dose buildup region or at material interfaces. Uncertainties due to improper detector positioning must also be quantified as these measurements are often performed in regions of high dose gradients.

When using modern IMRT techniques there can be strong perturbation effects or corrections needed with ion chamber and other detector measurements. However, as with other algorithms, detector perturbations are generally not accounted for in verification of MC-based treatment planning algorithms. MC methods may be used to model the physical details of the detectors to provide a better understanding of measurements performed under nonequilibrium conditions. These approaches are available to developers of MC techniques but further advancement and investigation of this technology is necessary before specific recommendations can be made. Measurements form the basis for benchmarking the accuracy of dose calculation algorithms in clinical radiotherapy. It is therefore important that the uncertainties in the measuring techniques be accounted for.

III.E.9. Example experimental tests

Example tests for verification of MC dose algorithms are provided in Table IV. These tests, intended as a supplement to those provided in the AAPM TG 53 report,[48] aim to evaluate the various components of the MC algorithm.

III.E.10. Timing issues

The issue of calculation time is of considerable importance in the clinical treatment planning process. To provide a perspective on this subject, users of commercial MC dose algorithms were asked to submit calculation times for photon and electron treatment plans. The treatment plans ranged in complexity from simple AP beams to dynamically delivered IMRT plans. These results are summarized in Table V. As noted in Table V, photon beam calculations with PEREGRINE are approximately a factor of 2–3 times slower than those for conventional planning systems. However, these times will be reduced using faster, currently available processors (>2 GHz versus 800 MHz). Electron beam calculations with VMC++ (Nucletron) are roughly equivalent in processing times to conventional systems. The timing comparisons provide good justification of the fact that significant processing times are no longer a concern for routine clinical MC dose calculation. MC calculations for a plan with one beam take the same time as that for a plan with multiple beams (of approximately the same field size), for the same statistical uncertainty to the PTV. This is because the statistical uncer-

TABLE IV. Partial listing of example specific tests, phantom designs, and detector measurements for Monte Carlo treatment planning systems. These tests are intended as a supplement to those detailed in the AAPM TG-53 (Ref. 48) and other related publications (Refs. 49 and 50) for algorithmic verification. Note that extreme care should be taken when performing many of these measurements as they are, in some instances, highly sensitive to the measurement setup conditions.

Test description	Reason	Phantom design
• Water depth doses and profiles—emphasis on large open field sizes, ($>30\times30$ cm^2) • 2D planar dose perpendicular to the beam cax for large open fields	To evaluate the beam model accuracy—test is sensitive to structures like the flattening filter and other parameters, such as the electron-on-target energy.	• Depth doses and profiles at multiple depths measured in a water phantom using a cylindrical ion chamber.• 2D planar dose at multiple depths in solid water using film.
• 2D planar dose perpendicular to the beam cax of large MLC-shaped fields (see Fig. A3–1 of the AAPM TG 53 report)[a] • Dose profiles under the closed MLC leaves, perpendicular to the direction of motion (see Fig. 7).[b]	To evaluate the accuracy of the MLC model, leaf-tip penumbra and leaf transmission.	• 2D planar dose at multiple depths in solid water using film. • Dose profiles under closed MLC leaves measured with film or small volume detector (diode, TLD, pinpoint chamber, diamond detector).
• Small field (1×1 cm^2–4×4 cm^2) depth doses in low density media; larger field sizes should also be tested. • Penumbral broadening; lateral dose spreading in lung assessed over a range of field sizes (2×2–30×30 cm^2).	To evaluate the transport algorithm accuracy—use of high energies (>10 MV) and low density media emphasizes electronic disequilibrium effects.	• Depth doses in a layered phantom (see Fig. 8 and Fig. 1 of Rice *et al.*)[c] consisting of solid water and low density material (lung equivalent or cork) measured with small volume detector (diode, TLD, pinpoint chamber, diamond detector, at multiple point depths) or with film. Beam is directed perpendicularly to the slabs. • 2D penumbral measurements with film in planes perpendicular to the beam cax at depths above, below, and within the low density slab in the layered phantom. Beam should also be directed parallel to the slabs to evaluate interface effects.[d]
• Depth doses in high density media over a range of field sizes, 3×3–30×30 cm^2.	To evaluate the transport algorithm accuracy in high density media, such as cortical-bone equivalent slabs.	• Depth doses in a layered phantom consisting of solid water and high density material (cortical bone equivalent) measured with small volume detector (diode, TLD, pinpoint chamber, diamond detector) for smaller field sizes, or with film.
• Point doses in the vicinity of tissue interfaces (tissue/lung and tissue/bone), over a range of field sizes, 3×3 –30×30 cm^2.	To evaluate the algorithmic accuracy in the perturbed dose field at tissue interfaces.	See, for example, Fig. 1 of Ref. 230. Dose measured with film or with small volume detector, where possible, at incremental depths, for example, 0.2, 0.5, 1.0, 2.0, and 5.0 cm anterior/posterior to the medial and proximal tissue/lung equivalent and tissue/bone-equivalent interfaces.
• Dose evaluation in clinical treatment planning, for simple, intermediate and complex static treatment plans as well as IMRT plans, in anthropomorphic phantoms.	To assess the accuracy of dose calculation to points located within structures of different densities and receiving different doses based on the treatment plan.	• Dose measured with small volume detectors within inserts of different materials, ranging from air to cortical bone-equivalent. Plans designed should include simple, intermediate, complex static and IMRT beam arrangements. Anthropomorphic phantoms should be CT-imaged for planning purposes.

[a]See Ref. 48.
[b]See Ref. 83.
[c]See Ref. 229.
[d]See Ref. 222.

tainty is determined by the number of particles passing through a volume, and this number can be held constant when performing a MC plan with multiple beams. This is a distinct advantage over conventional algorithms where computational time scales linearly with the number of beams.

IV. CLINICAL IMPLICATIONS OF MONTE CARLO-CALCULATED DOSE DISTRIBUTIONS

IV.A. Introduction

In spite of our confidence in the improved dose calculation accuracy with a suitably commissioned clinical MC algorithm, we are confronted with the following clinical question: What is the effect of more accurate MC dose distributions on patient clinical outcome? To answer this question, we will need to investigate the correlation of MC calculated dose distributions with clinical outcome (in terms of tumor control and normal tissue toxicity).

To date the evidence directly correlating the improved accuracy of MC-calculated dose distributions with clinical outcome is scant. Investigations by De Jaeger *et al.*,[237] (which included convolution but not MC calculations), Chetty *et al.*,[238] and Lindsay *et al.*[239] are among the first

TABLE V. Summary of timing results for clinical treatment plans from currently available commercial Monte Carlo systems. Data for photon beams were provided by author G.E., performed using the PEREGRINE (Nomos division, North American Scientific) system and those for electrons by author J.E.C., conducted with Nucletron. The (1σ) relative statistical uncertainty was approximately 2% in the maximum dose voxel for the Nomos calculations and roughly 1%–1.5% (in the average depth dose along the central axis) for the Nucletron electron beam calculations. Eclipse calculations were reported with an uncertainty of 1%–2% in the mean dose of all voxels receiving more than 50% of the maximum dose within the body of the contour. Readers should be cautioned that the timing results are subject to large uncertainties due to differences in compilers, memory size, cache size, etc.

Monte Carlo code/configuration	Description of treatment plan	Time estimate (min)
PEREGRINE (Nomos, North American Scientific) 16 processors (8-dual), Pentium III, 800 MHz	AP beam, 6 MV photons, 10×10 cm^2 in a water phantom, cubic voxels with 2 mm sides	48
	5 field, 6 MV CRT prostate plan, $\sim7\times11$ cm^2, cubic voxels with ~2.4 mm sides	89
	5 field, 6 MV prostate plan with modulation delivered with DMLC (exposed field $\sim4\times8$ cm^2), cubic voxels with ~2.4 mm sides	71
Masterplan (VMC++, Nucletron), single CPU Pentium IV XEON, 2.2 GHz	AP beam, 6 MeV electrons, 10×10 cm applicator, water phantom, cubic voxels with 4.9 mm sides	4.2
	AP beam, 17 MeV electrons, 10×10 cm applicator, water phantom, cubic voxels with 4.9 mm sides	8.2
	AP beam, 20 MeV electrons, 15×15 cm applicator, water phantom, cubic voxels with 3.9 mm sides	21
	Breast boost treatment plan: 2 fields, 11 MeV electrons, $\sim3\times3$ cm cutouts, cubic voxels with 5.4 mm sides	8.6 (both fields)
eMC (MMC, Eclipse, Varian) single CPU Pentium IV XEON, 2.4 GHz	AP beam, 6, 12, 18 MeV electrons, 10×10 cm applicator, water phantom, cubic voxels with 5.0 mm sides	~3, 4, 4 (6, 12, 18 MeV, respectively)

studies evaluating the influence of improved dose distributions on outcome observed in patients treated with lung cancer. The study by De Jaeger *et al.*,[237] in which lung cancer treatment plans were retrospectively recalculated using a convolution/superposition (CS)-based algorithm (initially calculated with an equivalent-path-length (EPL) algorithm), showed clinically significant differences between calculated and observed incidences of radiation pneumonitis. They demonstrated that the calculated incidence of radiation pneumonitis correlated better with observed incidence when using dose distributions calculated with CS rather than EPL algorithms.[237] Although this study was carried out using a CS algorithm, it provides strong support that the dose-response relationships determined with correction-based algorithms will be different than those computed with model-based methods.[237] With the sometimes large differences observed between the doses calculated with CS and MC algorithms,[240] the MC method is likely to add a higher degree of accuracy to the dose-effect relationships, and will be instrumental in putting these relationships on a more solid footing.

There is clearly a need for more studies addressing the clinical impact of MC-calculated dose distributions. The use of retrospective data may provide a useful means to perform such studies. Retrospective dose assessments of already existing local tumor control and normal tissue complications, using doses recalculated with MC algorithms, may give an early indication of the clinical utility of the MC method, and may also help physicians determine how to use the new MC-calculated doses.[206] Retrospective analyses should eventually show us how to make use of this information in a prospective way.[206]

IV.B. Clinical examples

In reviewing the literature on clinical treatment planning, one should keep in mind that the dose differences found between MC-based and conventional algorithms will be highly dependent on the beam arrangements, field sizes, beam energies, tumor size, and location. This is particularly true in anatomical sites where the target is situated near tissues with widely varying densities, such as the lung and head/neck. For example, due to electron transport issues, differences found in a lung CRT treatment plan using small field sizes and 15 MV photons may be much larger than those found with a standard AP/PA lung plan, using large field sizes and 6 MV photon beams. The reader should therefore be advised that, although there is a general consensus on the importance of the MC method in sites such as the lung, the dosimetry in many of the reported studies is based on specific conditions. The following literature review will focus on treatment planning in the lung and head and neck since differences between MC-based and conventional algorithms are likely to be smaller in other external beam treatment sites. This does not include the potential for improvement in dose estimates in any site from accurate simulation of beam modifiers, in particular, the MLC for delivering IMRT. More studies using MC-based dose calculation techniques in clinical treatment planning are warranted to better quantify the dosimetric and clinical benefits of these algorithms.

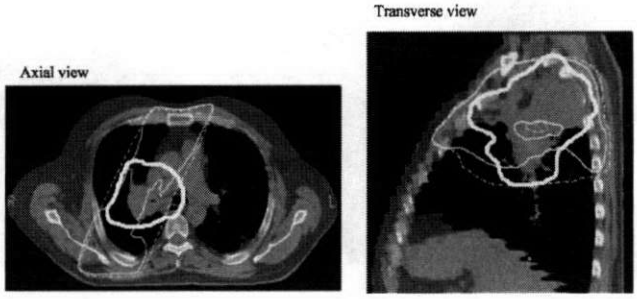


FIG. 11. Opposed, oblique field treatment plan (15 MV photons) showing the 100% isodose coverage for MC (modified DPM, University of Michigan/UMPlan) in the solid line, and an equivalent path length (EPL, University of Michigan/UMPlan) algorithm in the dashed line. The PTV is demarcated in white.

An excellent review of MC-based dose calculation methods in clinical planning for various treatment sites, including the breast and prostate is provided in the article by Reynaert *et al.*[241]

IV.B.1. Photon beam treatment planning

IV.B.1.a. Lung Perhaps the strongest motivation for the need for MC dose calculation comes from treatment planning for lung cancer. This is because electron transport issues, not accounted for accurately with conventional algorithms, are exacerbated in the low density tissues. A consequence of electronic disequilibrium in the lung is the underdosage of the PTV, as shown in Fig. 11. The penumbral widening in the dose distribution as a result of the increased electron scattering in the lung is illustrated for a conformal lung plan in Fig. 12. Depending on the location and size of the tumor, and the beam energy, underdosage of the PTV in lung planning may be significant. In addition to the target coverage, dose to normal tissues, particularly the normal lung, may be equally affected. Numerous lung planning studies have shown sometimes substantial differences (10% −20%) between conventional and MC algorithms.[117,190,192,214,216,242–246]

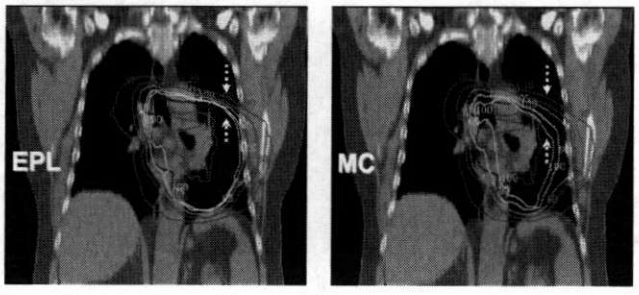


FIG. 12. Isodose distributions for a 3D conformal lung plan (15 MV photons) calculated using an equivalent path length (EPL, University of Michigan/UMPlan) algorithm on the left and a MC calculation (modified DPM, UMPlan) on the right. A distinct penumbral broadening in the MC-based dose distribution (on the right) is observed due to the increased electron scattering in the lung tissue. This effect is not as pronounced in the EPL-calculated dose distribution.

Due to the complicated dosimetric issues associated with treatment planning in the lung, more studies and comparisons using the MC method are encouraged. The utilization of advanced techniques, such as extracranial stereotactic radiotherapy using 10–24 Gy fractions,[247] for the treatment of early stage lung cancer, may increase the importance of accurate dosimetry and the clinical importance of MC calculated doses. One of the important next steps must be to evaluate how the improved MC dose distributions will clinically impact outcome for tumors and normal tissues.

IV.B.1.b. Head and neck There have been numerous studies on MC head and neck CRT and IMRT planning.[190,192,193,214,248–250] These studies have shown in general that dose differences between convolution and MC algorithms are dosimetrically insignificant. However, it has also been shown that dose differences for tumors located in the presence of air cavities can be significant due to inaccurate calculation of the photon/electron energy fluence inside and around air cavities using Batho and ETAR methods[214] as well as a collapsed cone convolution algorithm.[250,251]

IV.B.1.c. Other treatment sites Readers are encouraged to review articles published on MC-based treatment planning in other anatomical sites, which include: brain,[89,91,219,252–254] breast,[216,255] and prostate.[193,214,216,256,257]

IV.B.2. Electron beam treatment planning

There have been numerous studies on the use of MC calculations in electron beam treatment planning.[26,111,115,121,258–262] In some instances, significant differences between pencil beam dose distributions and MC calculations have been demonstrated, particularly in regions in or near air cavities and lung or bone tissues.[115,263] Large differences have also been observed for small irregular fields, beams with oblique incidence, and for extended SSD treatments.[115,263] The use of the MC method represents a significant improvement in accuracy for electron beam dose calculation compared with conventional algorithms. Electron MC calculations require far fewer primary histories to achieve a given uncertainty on the dose being calculated because electrons deposit their energy in a more continuous manner, in a smaller volume and mostly starting from the same location (the patient surface). This timing advantage has allowed the development of commercial electron beam MC algorithms which are time efficient (taking on the order of minutes per plan) even on a single processor.[26]

From phantom and routine patient planning calculations based on experience with a commercial electron beam MC algorithm (VMC++, currently Nucletron), Cygler *et al.*[26] have reported that MC-generated monitor unit (MU) values may differ from the homogeneous (water phantom) values by as much as 10%, depending on surface irregularities and inhomogeneities present in the irradiated volume. This situation is frequently encountered for head and neck sites, especially in the region of the nasal cavities. In such cases, surface irregularities and missing tissue lateral to the air cavity necessitate larger MUs to deliver the same prescribed dose versus flat surface anatomies. As is the case with photon beams,

it is important that the statistical uncertainties in the single voxel MU calculations be reported.

IV.C. Association of Monte Carlo calculated dose distributions with clinical outcome

As with other changes to the therapy treatment process, with the implementation of a new dose calculation algorithm such as the MC method, users should correlate doses and prescriptions with respect to previous clinical experience.

Chetty *et al.*[238] studied dose-effect relationships (for tumor and normal tissues) by recalculating dose distributions retrospectively using the MC method for patients treated on a nonsmall cell lung cancer (NSCLC) dose escalation protocol[264]—original plans were generated using an EPL algorithm. Follow-up data was evaluated from CT scans taken six months to two years postradiation therapy. Follow-up CT scans were fused with initial treatment planning scans and the original and replanned dose distributions were mapped onto the anatomy to establish associations between dose and regions of local recurrence and normal lung damage (radiation-induced pneumonitis).[238] Preliminary results of this study showed that the originally planned PTVs are sometimes significantly underdosed with MC calculations compared with the EPL algorithm.[238] For normal lung tissue, the correlation of dose with normal tissue complications was also found to differ, but also showed that beam model differences (not related to the dose calculation algorithm in the patient) are important and must be considered for an unbiased comparison of dose calculated by different algorithms.[136,238,241,265]

Lindsay *et al.*[239] performed retrospective MC-based recalculations of a large group of lung cancer treatment plans and showed significant differences in dose indices (V20, maximum lung dose and mean GTV dose) between plans without heterogeneity correction and MC calculations. Moreover, correlations between V20 and observed radiation pneumonitis in this study were found to be different between plans without heterogeneity correction and MC-based treatment plans.[239]

Despite the preliminary (and somewhat anecdotal) nature of the evidence thus far, observations suggest that more accurate dose calculations will reveal inadequate target coverage or hot spots in certain areas of organs at risk that could lead to differences in outcome. Complete MC recalculations of delivered dose distributions that include the effects of other factors, such as organ motion and patient setup errors, will be required to refine these correlation studies. In addition, it should be noted that the efficacy of radiation therapy is also dependent upon individual patient response.

The clinical evidence thus far, albeit preliminary and retrospective, provides support that dose delivery based on MC treatment plans, particularly for lung cancer, have the potential to result in clinically significant changes. Kong *et al.*[266] have shown that dose significantly impacts local control and overall survival for NSCLC; local control was found to increase at a rate of 1.3% per gray above the conventional dose fractionation scheme (63–69 Gy in 2.0 Gy fractions). This suggests that even small differences in dose distributions, as a result of inaccurate dose calculation, are likely to affect local control and survival for patients with NSCLC. Further studies of MC dose distributions in the lung and in other sites, such as the head/neck, are necessary and are encouraged in order to unequivocally evaluate the clinical utility of MC-calculated doses.

V. SUMMARY

We wish to reiterate that, from a clinical implementation standpoint, the MC method should be treated as would any conventional dose algorithm. Proper implementation will require the clinical physicist to understand, on some level, the fundamentals of the algorithm as well as the possible pitfalls associated with its clinical implementation, much as one would for any dose algorithm. In addition to providing an educational review of the MC dose calculation method, we have identified issues that will need to be considered by developers, vendors, and end users of MC-based techniques, ultimately to ensure that patient dose calculations are effected safely.

In conclusion, we present a summary of some of the issues which are important in the implementation and clinical use of MC dose calculation algorithms. The recommendations summarized here are not meant to be prescriptive; rather they are intended as a preliminary guide for medical physicists to ensure thoughtful and safe implementation of clinical MC algorithms. We anticipate that many of the areas of concern will be studied in further detail in future, more specific task group reports.

V.A. Treatment head simulation

(a) Vendors of MC-based dose calculation systems should be responsible for providing the necessary guidance and assistance with the beam modeling and benchmarking process. Such guidance includes the tuning of parameters such as the electron energy, or the adjustment of model parameters (in measurement-driven models) to ensure that the beam model meets the required specifications.

(b) In reporting statistical uncertainties in the calculated patient dose, vendors should properly account for latent variance in the beam model if the model is based on treatment head simulation. Further, if the latent variance is a significant contributor to the total variance the number of phase-space particles in the beam simulation should be increased.

V.B. Patient simulation

V.B.1. Statistical uncertainties

We recommend that quantities, such as $F_{D>0.5D_{max}}$, or F_{PTV}, or F_{PRV} for doses to the specific volumes, PTV or PRV respectively, be adopted as a standard method of reporting fractional statistical uncertainties in dose averaged over the relevant volume. The sole use of dose uncertainties to indi-

vidual voxels, such as the maximum dose voxel, D_{max}, should be avoided. Additionally, reporting of single voxel doses or doses over patient subvolumes should be accompanied by their respective statistical uncertainties. In situations where doses in individual voxels are important, such as D_{max} to a serial organ like the spinal cord, it may be necessary to simulate a large enough number of histories so that $s_{D_{max}}$ is very small. This will ensure that the absolute uncertainty in D_{max} will also be small. The required statistical precision required for individual voxel dose estimates should be decided upon with guidance from the clinical team.

V.B.2. Variance reduction techniques, efficiency enhancing methods, and other parameters

Users should understand the influence on the dose accuracy of variance reduction implementations and approximate methods used to improve the calculation efficiency, as well as any other accessible parameters of importance to the MC dose algorithm. Appropriate documentation on the methods used and their influence should be made available to the user. More studies on the influence of efficiency enhancing methods in clinical treatment planning are warranted, as it is likely that the tradeoffs between speed and accuracy will not be the same in different anatomical sites.[267] Where possible, vendors should provide users with the flexibility to adjust parameter inputs for these efficiency enhancement techniques but implement default values which are conservative, i.e., accurate in all situations.

V.B.3. Dose prescriptions

Vendors are strongly discouraged from using single voxel (point) doses for dose prescription and monitor unit calculations in their MC-based treatment planning systems. Rather, doses should be prescribed to volumes larger than a single voxel, such as the volume contained by an isodose surface. Current users of MC-based planning systems are encouraged to find ways of circumventing point-based dose prescriptions if their systems are not flexible enough to allow otherwise.

V.B.4. CT-to-material conversions

The use of conversion techniques based purely on mass density (i.e., assuming the only patient material is water, but with varying density), as employed in conventional algorithms, is discouraged with MC simulation because these methods ignore dependencies of particle interactions on the materials, which can lead to notable discrepancies in high atomic number materials. The conversions should include the use of both mass density and the atomic compositions of the materials. Appropriate documentation on the CT number-to-material conversion method used by the software should be accessible to the user.

V.B.5. Dose-to-water and dose-to-medium

MC dose results should: (a) explicitly indicate the material to which the dose is computed, (b) allow conversion between D_m and D_w using the methods discussed in Sec. III D 4, or other methods as developed in future investigations. It is strongly encouraged that appropriate documentation on the dose-to-water conversion method used by the software be provided to the user.

V.C. Experimental verification

V.C.1. Examples of specific tests

In addition to the standard testing necessary for conventional dose algorithms, it is strongly recommended that additional tests (e.g., as suggested in Sec. III E 3) be included to evaluate the accuracy of MC calculations under situations where the MC method is known to perform better than conventional algorithms.

V.C.2. Verification calculations

MC verification calculations should be performed under the same conditions as the experiments. Phantoms used should be CT-imaged for planning purposes. Statistical uncertainties for verification plans should be reported and included in the comparison of calculations with measurements. More studies of the influence of detector perturbations using MC calculations are encouraged.

V.C.3. Measurement uncertainties

It is important for the user to understand the limitations of the various measurement devices as they are used in different situations. It is recommended that realistic measurement uncertainties be assessed and included when evaluating MC verification calculations against measurements.

Although the implementation of MC treatment planning will require clinical physicists to once again understand a new technology, one should view this in a positive light. Properly implemented MC algorithms will provide dose calculation with sufficient accuracy (in at least a single instance of the patient geometry) such that dose calculation is no longer a source of meaningful uncertainty in the radiotherapy planning process. In addition, although the details are sometimes complex, the underlying idea of simulating the actual transport of individual particles is simpler to understand than many other algorithms because it is based on actual physical processes familiar to the medical physicist. Finally, this may well be the last new dose calculation algorithm that medical physicists will need to learn since MC techniques provide the highest level of accuracy for dose calculations in radiotherapy treatment planning.

ACKNOWLEDGMENTS

I.J.C. acknowledges the support of his colleagues, in particular, Dick Fraass, Daniel McShan, and Randy Ten Haken, at the University of Michigan, Department of Radiation Oncology. Thanks to Neelam Tyagi and Mihaela Rosu for their help with compiling references. We are grateful to Lesley Buckley, Dan La Russa, and Randy Taylor of Carleton University for commenting on early versions of the manuscript. I.J.C. is thankful to Alex Bielajew for helpful discussions on various topics in this report. I.J.C. acknowledges the support

of NIH Grant Nos. R01CA106770 (UNMC) and P01CA59827 (previously at the University of Michigan). D.W.O.R. is supported by the Canada Research Chairs program and the Natural Sciences and Engineering Research Council of Canada. B.A.F. acknowledges the support of NIH Grant No. R01CA104777. J.V.S. acknowledges the support of NIH Grant No. R01CA98524. J. S. is a research scientist of the National Cancer Institute Canada supported with funds from the Canadian Cancer Society. Finally, we are indebted to Alan Nahum for his careful review of this report which helped us improve the pedagogical aspects and other topics significantly.

[a]Electronic mail: ichetty@unmc.edu

[1]N. Papanikolaou, J. Battista, A. Boyer, C. Kappas, E. Klein, T. Mackie, M. Sharpe, and J. Van Dyk, "AAPM Report No. 85: Tissue inhomogeneity corrections for megavoltage photon beams," in *AAPM Report No. 85* (Medical Physics, Madison, WI, 2004), pp. 1–135.

[2]A. Dutreix, "When and how can we improve precision in radiotherapy?," Radiother. Oncol. **2**, 275–292 (1984).

[3]C. G. Orton, P. M. Mondalek, J. T. Spicka, D. S. Herron, and L. I. Andres, "Lung corrections in photon beam treatment planning: Are we ready?," Int. J. Radiat. Oncol. Biol. Phys. **10**, 2191–2199 (1984).

[4]J. G. Stewart and A. W. Jackson, "The steepness of the dose response curve both for tumor cure and normal tissue injury," Laryngoscope **85**, 1107–1111 (1975).

[5]M. Goitein and J. Busse, "Immobilization error: Some theoretical considerations," Radiology **117**, 407–412 (1975).

[6]http://eom.springer.de/B/b017750.htm

[7]N. Metropolis, "The beginning of the MC Method," Los Alamos Sci. **15**, 125–130 (1987); see http://library.lanl.gov/la-pubs/00326866.pdf.

[8]D. W. O. Rogers and A. F. Bielajew, in *The Dosimetry of Ionizing Radiation*, edited by B. Bjarngard, K. Kase, and F. Attix (Academic, New York, 1990), Vol. III, pp. 427–539.

[9]M. J. Berger, in *Methods in Computational Physics*, edited by S. Fernbach, B. Alder, and M. Rothenberg (Academic, New York, 1963), Vol. 1.

[10]M. Berger and S. Seltzer, "ETRAN Monte Carlo code system for electron and photon transport through extended media," Radiation Shielding Information Center (RSIC) Report CCC-107, Oak Ridge National Laboratory, Oak Ridge, TN, 1973.

[11]J. F. Briesmeister, "MCNP—A general Monte Carlo *N* -particle transport code, version 4A," Report LA-12625-M, Los Alamos National Laboratory, Los Alamos, NM, 1993.

[12]W. R. Nelson, H. Hirayama, and D. W. O. Rogers, "The EGS4 code system," Report SLAC-265, Stanford Linear Accelerator, Stanford, CA, 1985.

[13]P. L. Petti, M. S. Goodman, T. A. Gabriel, and R. Mohan, "Investigation of buildup dose from electron contamination of clinical photon beams," Med. Phys. **10**, 18–24 (1983).

[14]R. Mohan, C. Chui, and L. Lidofsky, "Energy and angular distributions of photons from medical linear accelerators," Med. Phys. **12**, 592–597 (1985).

[15]M. Udale, "A Monte Carlo investigation of surface doses for broad electron beams," Phys. Med. Biol. **33**, 939–954 (1988).

[16]R. L. Ford and W. R. Nelson, "The EGS code system—Version 3," Report SLAC-210, Stanford Linear Accelerator, Stanford, CA, 1978.

[17]I. Kawrakow and D. W. O. Rogers, "The EGSnrc code system: Monte Carlo simulation of electron and photon transport," Technical Report PIRS-701, National Research Council of Canada, Ottawa, Ontario, 2000.

[18]P. R. Almond, P. J. Biggs, B. M. Coursey, W. F. Hanson, M. S. Huq, R. Nath, and D. W. O. Rogers, "AAPM's TG-51 protocol for clinical reference dosimetry of high-energy photon and electron beams," Med. Phys. **26**, 1847–1870 (1999).

[19]B. A. Faddegon, P. O'Brien, and D. L. Mason, "The flatness of Siemens linear accelerator x-ray fields," Med. Phys. **26**, 220–228 (1999).

[20]M. B. Tacke, H. Szymanowski, U. Oelfke, C. Schulze, S. Nuss, E. Wehrwein, and S. Leidenberger, "Assessment of a new multileaf collimator concept using GEANT4 Monte Carlo simulations," Med. Phys. **33**, 1125–1132 (2006).

[21]A. Ito, in *Monte Carlo Transport of Electrons and Photons*, edited by W. R. Nelson, T. M. Jenkins, A. Rindi, A. E. Nahum, and D. W. O. Rogers (Plenum, New York, 1988), pp. 573–598.

[22]A. E. Nahum, in *Monte Carlo Transport of Electrons and Photons*, edited by W. R. Nelson, T. M. Jenkins, A. Rindi, A. E. Nahum, and D. W. O. Rogers (Plenum, New York, 1988), pp. 3–20.

[23]D. W. O. Rogers, B. A. Faddegon, G. X. Ding, C. M. Ma, J. We, and T. R. Mackie, "BEAM: A Monte Carlo code to simulate radiotherapy treatment units," Med. Phys. **22**, 503–524 (1995).

[24]C.-M. Ma and S. B. Jiang, "Monte Carlo modelling of electron beams from medical accelerators," Phys. Med. Biol. **44**, R157–R189 (1999).

[25]F. Verhaegen and J. Seuntjens, "Monte Carlo modelling of external radiotherapy photon beams," Phys. Med. Biol. **48**, R107–R164 (2003).

[26]J. E. Cygler, G. M. Daskalov, G. H. Chan, and G. X. Ding, "Evaluation of the first commercial Monte Carlo dose calculation engine for electron beam treatment planning," Med. Phys. **31**, 142–153 (2004).

[27]C. L. Hartmann Siantar, "Description and dosimetric verification of the PEREGRINE Monte Carlo dose calculation system for photon beams incident on a water phantom," Med. Phys. **28**, 1322–1337 (2001).

[28]P. Andreo, "Monte Carlo techniques in medical radiation physics," Phys. Med. Biol. **36**, 861–920 (1991).

[29]D. E. Raeside, "Monte Carlo principles and applications," Phys. Med. Biol. **21**, 181–197 (1976).

[30]D. W. O. Rogers, "Fifty years of Monte Carlo simulations for medical physics," Phys. Med. Biol. **51**, R287–R301 (2006).

[31]J. E. Turner, H. A. Wright, and R. N. Hamm, "A Monte Carlo primer for health physicists," Health Phys. **48**, 717–733 (1985).

[32]F. B. Brown, "MCNP—A general Monte Carlo-particle transport code, version 5," Report LA-UR-03 1987, Los Alamos National Laboratory, Los Alamos, NM, 2003.

[33]S. Agostinelli, "GEANT4—A simulation toolkit," Nucl. Instrum. Methods Phys. Res. A **506**, 250–303 (2003).

[34]J. Baro, J. Sempau, J. M. Fernandez-Varea, and F. Salvat, "PENELOPE—An algorithm for Monte-Carlo simulation of the penetration and energy-loss of electrons and positrons in matter," Nucl. Instrum. Methods Phys. Res. A **100**, 31–46 (1995).

[35]M. Fippel, "Fast Monte Carlo dose calculation for photon beams based on the VMC electron algorithm," Med. Phys. **26**, 1466–1475 (1999).

[36]I. Kawrakow, "VMC++ , electron and photon Monte Carlo calculations optimized for radiation treatment planning," in *Advanced Monte Carlo for Radiation Physics, Particle Transport Simulation and Applications: Proceedings of the Monte Carlo 2000 Meeting Lisbon*, edited by A. Kling, F. Barao, M. Nakagawa, L. Tavora, and P. Vaz (Springer, Berlin, 2001), pp. 229–236.

[37]I. Kawrakow, M. Fippel, and K. Friedrich, "3D electron dose calculation using a Voxel based Monte Carlo algorithm (VMC)," Med. Phys. **23**, 445–457 (1996).

[38]H. Neuenschwander and E. J. Born, "A macro Monte-Carlo method for electron-beam dose calculations," Phys. Med. Biol. **37**, 107–125 (1992).

[39]J. Sempau, S. J. Wilderman, and A. F. Bielajew, "DPM, a fast, accurate Monte Carlo code optimized for photon and electron radiotherapy treatment planning dose calculations," Phys. Med. Biol. **45**, 2263–2291 (2000).

[40]I. Kawrakow, "Accurate condensed history Monte Carlo simulation of electron transport. I. EGSnrc, the new EGS4 version," Med. Phys. **27**, 485–498 (2000).

[41]A. F. Bielajew and D. W. O. Rogers, in *Monte Carlo Transport of Electrons and Photons*, edited by W. R. Nelson, T. M. Jenkins, A. Rindi, A. E. Nahum, and D. W. O. Rogers (Plenum, New York, 1988), pp. 115–137.

[42]I. Kawrakow and A. F. Bielajew, "On the condensed history technique for electron transport," Nucl. Instrum. Methods Phys. Res. B **142**, 253–280 (1998).

[43]E. W. Larsen, "A theoretical derivation of the condensed history algorithm," Ann. Nucl. Energy **19**, 701–714 (1992).

[44]I. Kawrakow, "Accurate condensed history Monte Carlo simulation of electron transport. II. Application to ion chamber response simulations," Med. Phys. **27**, 499–513 (2000).

[45]I. Kawrakow and M. Fippel, "Investigation of variance reduction techniques for Monte Carlo photon dose calculation using XVMC," Phys. Med. Biol. **45**, 2163–2183 (2000).

[46]I. Kawrakow, D. W. O. Rogers, and B. R. B. Walters, "Large efficiency improvements in BEAMnrc using directional bremsstrahlung splitting," Med. Phys. **31**, 2883–2898 (2004).

[47]I. Kawrakow and B. R. B. Walters, "Efficient photon beam dose calculations using DOSXYZnrc with BEAMnrc," Med. Phys. **33**, 3046–3056 (2006).
[48]B. Fraass, K. Doppke, M. Hunt, G. Kutcher, G. Starkschall, R. Stern, and J. Van Dyke, "American Association of Physicists in Medicine Radiation Therapy Committee Task Group 53: Quality assurance for clinical radiotherapy treatment planning," Med. Phys. **25**, 1773–1829 (1998).
[49]IAEA-Technical Report Series No. 430: Commissioning and quality qssurance of computerized planning systems for radiation treatment of cancer," in International Atomic Energy Agency, Vienna, 2004.
[50]J. Van Dyk, R. B. Barnett, J. E. Cygler, and P. C. Shragge, "Commissioning and quality assurance of treatment planning computers," Int. J. Radiat. Oncol. Biol. Phys. **26**, 261–273 (1993).
[51]D. W. O. Rogers and A. F. Bielajew, in *Monte Carlo Transport of Electrons and Photons*, edited by W. R. Nelson, T. M. Jenkins, A. Rindi, A. E. Nahum, and D. W. O. Rogers (Plenum, New York, 1988), pp. 407–419.
[52]I. Kawrakow, "On the efficiency of photon beam treatment head simulations," Med. Phys. **32**, 2320–2326 (2005).
[53]C.-M. Ma *et al.*, "A Monte Carlo dose calculation tool for radiotherapy treatment planning," Phys. Med. Biol. **47**, 1671–1689 (2002).
[54]J. A. Halbleib, in *Monte Carlo Transport of Electrons and Photons*, edited by W. R. Nelson, T. M. Jenkins, A. Rindi, A. E. Nahum, and D. W. O. Rogers (Plenum, New York, 1988), pp. 249–262.
[55]J. A. Halbleib and T. A. Melhorn, "ITS: The integrated TIGER series of coupled electron/photon Monte Carlo transport codes," Sandia Report SAND84–0573, Sandia National Laboratory, Albuquerque, NM, 1984.
[56]S. M. Seltzer, in *Monte Carlo Transport of Electrons and Photons*, edited by W. R. Nelson, T. M. Jenkins, A. Rindi, A. E. Nahum, and D. W. O. Rogers (Plenum, New York, 1988), pp. 153–182.
[57]D. W. O. Rogers, B. Walters, and I. Kawrakow, "BEAMnrc Users Manual," NRC Report PIRS 509(a)revH, 2004.
[58]B. R. B. Walters and D. W. O. Rogers, "DOSXYZnrc Users Manual," NRC Report PIRS 794 (rev B), 2004.
[59]D. W. O. Rogers, "The role of Monte-Carlo simulation of electron-transport in radiation-dosimetry," Appl. Radiat. Isot. **42**, 965–974 (1991).
[60]T. R. Mackie, S. S. Kubsad, D. W. O. Rogers, and A. F. Bielajew, "The OMEGA project: Electron dose planning using Monte Carlo simulation," Med. Phys. **17**, 730 (abstract) (1990).
[61]J. Siebers and R. Mohan, "Monte Carlo and IMRT," in *Intensity Modulated Radiation Therapy, The State of the Art, Proceedings of the 2003 AAPM Summer School*, edited by T. R. Mackie and J. R. Palta (Advanced Medical, Madison, WI, 2003), pp. 531–560.
[62]A. Leal, F. Sanchez-Doblado, R. Arrans, M. Perucha, M. Rincon, E. Carrasco, and C. Bernal, "Monte Carlo simulation of complex radiotherapy treatments," Comput. Sci. Eng. **6**, 60–68 (2004).
[63]N. Tyagi, A. Bose, and I. J. Chetty, "Implementation of the DPM Monte Carlo Code on a parallel architecture for treatment planning applications," Med. Phys. **31**, 2721–2725 (2004).
[64]J. M. Fernandez-Varea, R. Mayol, J. Baro, and F. Salvat, "On the theory and simulation of multiple elastic-scattering of electrons," Nucl. Instrum. Methods Phys. Res. B **73**, 447–473 (1993).
[65]A. E. Schach von Wittenau, L. J. Cox, P. M. Bergstrom, Jr., W. P. Chandler, C. L. Hartmann Siantar, and R. Mohan, "Correlated histogram representation of Monte Carlo derived medical accelerator photon-output phase space," Med. Phys. **26**, 1196–1211 (1999).
[66]A. E. Schach von Wittenau, P. M. Bergstrom, Jr., and L. J. Cox, "Patient-dependent beam-modifier physics in Monte Carlo photon dose calculations," Med. Phys. **27**, 935–947 (2000).
[67]I. Kawrakow, "Electron transport: Multiple and plural scattering," Nucl. Instrum. Methods Phys. Res. B **108**, 23–34 (1996).
[68]I. Kawrakow and A. F. Bielajew, "Recent improvements and accuracy tests of the VOXEL Monte Carlo algorithm," Med. Phys. **24**, 1049 (abstract) (1997).
[69]P. J. Keall and P. W. Hoban, "Superposition dose calculation incorporating Monte Carlo generated electron track kernels," Med. Phys. **23**, 479–485 (1996).
[70]L. Wang, C. S. Chui, and M. Lovelock, "A patient-specific Monte Carlo dose-calculation method for photon beams," Med. Phys. **25**, 867–878 (1998).
[71]R. Doucet, M. Olivares, F. DeBlois, E. B. Podgorsak, I. Kawrakow, and J. Seuntjens, "Comparison of measured and Monte Carlo calculated dose distributions in inhomogeneous phantoms in clinical electron beams," Phys. Med. Biol. **48**, 2339–2354 (2003).
[72]M. Fippel, "Efficient particle transport simulation through beam modulating devices for Monte Carlo treatment planning," Med. Phys. **31**, 1235–1242 (2004).
[73]M. Fippel, F. Haryanto, O. Dohm, F. Nusslin, and S. Kriesen, "A virtual photon energy fluence model for Monte Carlo dose calculation," Med. Phys. **30**, 301–311 (2003).
[74]M. Fippel, I. Kawrakow, and K. Friedrich, "Electron beam dose calculations with the VMC algorithm and the verification data of the NCI working group," Phys. Med. Biol. **42**, 501–520 (1997).
[75]M. Fippel, W. Laub, B. Huber, and F. Nusslin, "Experimental investigation of a fast Monte Carlo photon beam dose calculation algorithm," Phys. Med. Biol. **44**, 3039–3054 (1999).
[76]M. Fippel and F. Nusslin, "Evaluation of a clinical Monte Carlo dose calculation code based on the ICCR benchmark test," Med. Phys. **28**, 1198 (abstract) (2001).
[77]I. Kawrakow and A. F. Bielajew, "On the representation of electron multiple elastic-scattering distributions for Monte Carlo calculations," Nucl. Instrum. Methods Phys. Res. B **134**, 325–336 (1998).
[78]H. Neuenschwander, T. R. Mackie, and P. J. Reckwerdt, "MMC—A high-performance Monte Carlo code for electron beam treatment planning," Phys. Med. Biol. **40**, 543–574 (1995).
[79]C. Cris, E. Born, R. Mini, H. Neuenschwander, and W. Volken, "A scaling method for multiple source models," in *Proceedings of the 13th ICCR*, edited by T. Bortfeld and W. Schlegel (Springer-Verlag, Heidelberg, 2000), pp. 411–413.
[80]P. Pemler, J. Besserer, U. Schneider, and H. Neuenschwander, "Evaluation of a commercial electron treatment planning system based on Monte Carlo techniques (eMC)," Z. Med. Phys. **16**, 313–329 (2006).
[81]C.-M. Ma, J. S. Li, T. Pawlicki, S. B. Jiang, and J. Deng, "MCDOSE—A Monte Carlo dose calculation tool for radiation therapy treatment planning," in *Proceedings of the 13th ICCR*, edited by T. Bortfeld and W. Schlegel (Springer-Verlag, Heidelberg, 2000), pp. 411–413.
[82]J. V. Siebers, P. J. Keall, J. Kim, and R. Mohan, "Performance benchmarks of the MCV Monte Carlo System," in *Proceedings of the 13th ICCR*, edited by T. Bortfeld and W. Schlegel (Springer-Verlag, Heidelberg, 2000), pp. 129–131.
[83]J. V. Siebers, P. J. Keall, J. O. Kim, and R. Mohan, "A method for photon beam Monte Carlo multileaf collimator particle transport," Phys. Med. Biol. **47**, 3225–3249 (2002).
[84]I. J. Chetty, P. M. Charland, N. Tyagi, D. L. McShan, B. A. Fraass, and A. F. Bielajew, "Photon beam relative dose validation of the DPM Monte Carlo code in lung-equivalent media," Med. Phys. **30**, 563–573 (2003).
[85]I. J. Chetty, N. Tyagi, M. Rosu, P. M. Charland, D. L. McShan, R. K. Ten Haken, B. A. Fraass, and A. F. Bielajew, "Clinical implementation, validation and use of the DPM Monte Carlo code for radiotherapy treatment planning," in *Nuclear Mathematical and Computational Sciences: A Century in Review, A Century Anew, Gatlinburg, TN* (American Nuclear Society, LaGrange Park, IL, 2003), Vol. 119, pp. 1–17.
[86]J. J. DeMarco, T. D. Solberg, and J. B. Smathers, "A CT-based Monte Carlo simulation tool for dosimetry planning and analysis," Med. Phys. **25**, 1–11 (1998).
[87]R. F. Aaronson, J. J. DeMarco, I. J. Chetty, and T. D. Solberg, "A Monte Carlo based phase space model for quality assurance of intensity modulated radiotherapy incorporating leaf specific characteristics," Med. Phys. **29**, 2952–2958 (2002).
[88]I. Chetty, J. J. DeMarco, and T. D. Solberg, "A virtual source model for Monte Carlo modeling of arbitrary intensity distributions," Med. Phys. **27**, 166–172 (2000).
[89]I. J. Chetty, J. J. DeMarco, T. D. Solberg, A. R. Arellano, R. Fogg, and A. V. Mesa, "A phase space model for simulating arbitrary intensity distributions for shaped radiosurgery beams using the Monte Carlo method," Radiosurgery **3**, 41–52 (2000).
[90]T. D. Solberg *et al.*, "A review of radiation dosimetry applications using the MCNP Monte Carlo code," Radiochim. Acta **89**, 337–355 (2001).
[91]T. D. Solberg, J. J. DeMarco, F. E. Holly, J. B. Smathers, and A. A. F. DeSalles, "Monte Carlo treatment planning for stereotactic radiosurgery," Radiother. Oncol. **49**, 73–84 (1998).
[92]M. K. Fix, P. Manser, E. J. Born, R. Mini, and P. Ruegsegger, "Monte Carlo simulation of a dynamic MLC based on a multiple source model," Phys. Med. Biol. **46**, 3241–3257 (2001).
[93]M. K. Fix, M. Stampanoni, P. Manser, E. J. Born, R. Mini, and P. Ruegsegger, "A multiple source model for 6 MV photon beam dose calculations using Monte Carlo," Phys. Med. Biol. **46**, 1407–1427 (2001).

[94]E. Poon and F. Verhaegen, "Accuracy of the photon and electron physics in GEANT4 for radiotherapy applications," Med. Phys. **32**, 1696–1711 (2005).
[95]J. Sempau, A. Sanchez-Reyes, and F. Salvat, "H. O. ben Tahar, S. B. Jiang, and J. M. Fernandez-Varea, "Monte Carlo simulation of electron beams from an accelerator head using PENELOPE," Phys. Med. Biol. **46**, 1163–1186 (2001).
[96]O. Chibani and X. A. Li, "Monte Carlo dose calculations in homogeneous media and at interfaces: A comparison between GEPTS, EGSnrc, MCNP, and measurements," Med. Phys. **29**, 835–847 (2002).
[97]O. Chibani and C. M. Ma, "Electron depth dose distributions in water, iron and lead: The GEPTS system," Nucl. Instrum. Methods Phys. Res. B **101**, 357–378 (1995).
[98]W. van der Zee, A. Hogenbirk, and S. C. van der Marck, "ORANGE: A Monte Carlo dose engine for radiotherapy," Phys. Med. Biol. **50**, 625–641 (2005).
[99]D. W. O. Rogers and R. Mohan, "Questions for comparisons of clinical Monte Carlo codes," in *Proceedings of the 13th ICCR*, edited by T. Bortfeld and W. Schlegel (Springer-Verlag, Heidelberg, 2000), pp. 120–122.
[100]B. A. Faddegon and I. Blevis, "Electron spectra derived from depth dose distributions," Med. Phys. **27**, 514–526 (2000).
[101]M. R. Bieda, J. A. Antolak, and K. R. Hogstrom, "The effect of scattering foil parameters on electron-beam Monte Carlo calculations," Med. Phys. **28**, 2527–2534 (2001).
[102]D. Sheikh-Bagheri and D. W. O. Rogers, "Sensitivity of megavoltage photon beam Monte Carlo simulations to electron beam and other parameters," Med. Phys. **29**, 379–390 (2002).
[103]A. Tzedakis, J. E. Damilakis, M. Mazonakis, J. Stratakis, H. Varveris, and N. Gourtsoyiannis, "Influence of initial electron beam parameters on Monte Carlo calculated absorbed dose distributions for radiotherapy photon beams," Med. Phys. **31**, 907–913 (2004).
[104]I. J. Chetty, P. M. Charland, N. Tyagi, D. L. McShan, B. Fraass, and A. F. Bielajew, "Experimental validation of the DPM Monte Carlo code for photon beam dose calculations in inhomogeneous media," Med. Phys. **29**, 1351 (abstract) (2002).
[105]B. Libby, J. Siebers, and R. Mohan, "Validation of Monte Carlo generated phase-space descriptions of medical linear accelerators," Med. Phys. **26**, 1476–1483 (1999).
[106]C. Bramoulle, F. Husson, and J. P. Manens, "Monte Carlo (PENELOPE code) study of the x-ray beams from SL Linacs (Elekta)," Phys. Med. **16**, 107–115 (2000).
[107]E. L. Chaney, T. J. Cullip, and T. A. Gabriel, "A Monte Carlo study of accelerator head scatter," Med. Phys. **21**, 1383–1390 (1994).
[108]G. X. Ding, "Energy spectra, angular spread, fluence profiles and dose distributions of 6 and 18 MV photon beams: Results of Monte Carlo simulations for a Varian 2100EX accelerator," Phys. Med. Biol. **47**, 1025–1046 (2002).
[109]R. A. C. Siochi, "Requirements for manufacturer supplied data for Monte Carlo simulation," in *Proceedings of the 15th International Conference on the Applications of Accelerators in Research and Industry* (The American Institute of Physics, Melville, 1999), pp. 1060–1065.
[110]G. G. Zhang, D. W. O. Rogers, J. E. Cygler, and T. R. Mackie, "Monte Carlo investigation of electron beam output factors versus size of square cutout," Med. Phys. **26**, 743–750 (1999).
[111]J. A. Antolak, M. R. Bieda, and K. R. Hogstrom, "Using Monte Carlo methods to commission electron beams: A feasibility study," Med. Phys. **29**, 771–786 (2002).
[112]B. Faddegon, E. Schreiber, and X. Ding, "Monte Carlo simulation of large electron fields," Phys. Med. Biol. **50**, 741–753 (2005).
[113]E. C. Schreiber and B. A. Faddegon, "Sensitivity of large-field electron beams to variations in a Monte Carlo accelerator model," Phys. Med. Biol. **50**, 769–778 (2005).
[114]C.-M. Ma, B. A. Faddegon, D. W. O. Rogers, and T. R. Mackie, "Accurate characterization of Monte Carlo calculated electron beams for radiotherapy," Med. Phys. **24**, 401–416 (1997).
[115]C.-M. Ma, E. Mok, A. Kapur, T. Pawlicki, D. Findley, S. Brain, K. Forster, and A. L. Boyer, "Clinical implementation of a Monte Carlo treatment planning system," Med. Phys. **26**, 2133–2143 (1999).
[116]C.-M. Ma and D. W. O. Rogers, "BEAM Characterization: A Multiple-Source Model," NRC Report PIRS-0509(C), 1995.
[117]C.-M. Ma, "Characterization of computer simulated radiotherapy beams for Monte-Carlo treatment planning," Radiat. Phys. Chem. **53**, 329–344 (1998).
[118]J. Deng, S. B. Jiang, A. Kapur, J. Li, T. Pawlicki, and C. M. Ma, "Photon beam characterization and modelling for Monte Carlo treatment planning," Phys. Med. Biol. **45**, 411–427 (2000).
[119]B. Faddegon, J. Balogh, R. Mackenzie, and D. Scora, "Clinical considerations of Monte Carlo for electron radiotherapy treatment planning," Radiat. Phys. Chem. **53**, 217–227 (1998).
[120]M. K. Fix, H. Keller, P. Ruegsegger, and E. J. Born, "Simple beam models for Monte Carlo photon beam dose calculations in radiotherapy," Med. Phys. **27**, 2739–2747 (2000).
[121]S. B. Jiang, A. Kapur, and C. M. Ma, "Electron beam modeling and commissioning for Monte Carlo treatment planning," Med. Phys. **27**, 180–191 (2000).
[122]S. B. Jiang, J. Deng, J. Li, P. Pawlicki, A. Boyer, and C.-M. Ma, "Modeling and commissioning of clinical photon beams for Monte Carlo treatment planning," in *Proceedings of the 13th ICCR*, edited by T. Bortfeld and W. Schlegel (Springer-Verlag, Heidelberg, 2000), pp. 434–436.
[123]J. Deng, S. B. Jiang, P. Pawlicki, J. Li, and C.-M. Ma, "Electron beam commissioning for Monte Carlo dose calculation," in *Proceedings of the 13th ICCR*, edited by T. Bortfeld and W. Schlegel (Springer-Verlag, Heidelberg, 2000), pp. 431–433.
[124]J. S. Li *et al.*, "Source modeling and beam commissioning for Siemens photon beams," Med. Phys. **29**, 1230 (abstract) (2002).
[125]J. Yang, J. S. Li, L. Qin, W. Xiong, and C. M. Ma, "Modelling of electron contamination in clinical photon beams for Monte Carlo dose calculation," Phys. Med. Biol. **49**, 2657–2673 (2004).
[126]J. Deng, T. Guerrero, C. M. Ma, and R. Nath, "Modelling 6 MV photon beams of a stereotactic radiosurgery system for Monte Carlo treatment planning," Phys. Med. Biol. **49**, 1689–1704 (2004).
[127]J. Deng, S. B. Jiang, T. Pawlicki, J. Li, and C. M. Ma, "Derivation of electron and photon energy spectra from electron beam central axis depth dose curves," Phys. Med. Biol. **46**, 1429–1449 (2001).
[128]J. J. Janssen, E. W. Korevaar, L. J. van Battum, P. R. Storchi, and H. Huizenga, "A model to determine the initial phase space of a clinical electron beam from measured beam data," Phys. Med. Biol. **46**, 269–286 (2001).
[129]S. Siljamaki, L. Tillikainen, H. Helminen, and J. Pyyry, "Determining parameters for a multiple-source model of a linear accelerator using optimization techniques," Med. Phys. **32**, 2113 (abstract) (2005).
[130]W. Ulmer, J. Pyyry, and W. Kaissl, "A 3D photon superposition/convolution algorithm and its foundation on results of Monte Carlo calculations," Phys. Med. Biol. **50**, 1767–1790 (2005).
[131]K. Aljarrah, G. C. Sharp, T. Neicu, and S. B. Jiang, "Determination of the initial beam parameters in Monte Carlo linac simulation," Med. Phys. **33**, 850–858 (2006).
[132]A. Ahnesjo and P. Andreo, "Determination of effective bremsstrahlung spectra and electron contamination for photon dose calculations," Phys. Med. Biol. **34**, 1451–1464 (1989).
[133]A. Ahnesjo and A. Trepp, "Acquisition of the effective lateral energy fluence distribution for photon beam dose calculations by convolution models," Phys. Med. Biol. **36**, 973–985 (1991).
[134]A. Ahnesjo, L. Weber, A. Murman, M. Saxner, I. Thorslund, and E. Traneus, "Beam modeling and verification of a photon beam multisource model," Med. Phys. **32**, 1722–1737 (2005).
[135]A. Catala, P. Francois, J. Bonnet, and C. Scouarnec, "Reconstruction of 12 MV bremsstrahlung spectra from measured transmission data by direct resolution of the numeric system $AF=T$," Med. Phys. **22**, 3–10 (1995).
[136]P. M. Charland, I. J. Chetty, L. D. Paniak, B. P. Bednarz, and B. A. Fraass, "Enhanced spectral discrimination through the exploitation of interface effects in photon dose data," Med. Phys. **31**, 264–276 (2004).
[137]E. Heath and J. Seuntjens, "Development and validation of a BEAMnrc component module for accurate Monte Carlo modelling of the Varian dynamic millennium multileaf collimator," Phys. Med. Biol. **48**, 4045–4063 (2003).
[138]H. H. Liu, F. Verhaegen, and L. Dong, "A method of simulating dynamic multileaf collimators using Monte Carlo techniques for intensity-modulated radiation therapy," Phys. Med. Biol. **46**, 2283–2298 (2001).
[139]N. Tyagi, J. M. Moran, D. W. Litzenberg, A. F. Bielajew, B. A. Fraass, and I. J. Chetty, "Experimental verification of a Monte Carlo-based MLC simulation model for IMRT dose calculation," Med. Phys. **34**, 651–663 (2007).
[140]T. C. Zhu *et al.*, "Output ratios in air: Report of the AAPM Task Group No. 74," Med. Phys. (submitted).
[141]T. C. Zhu and B. E. Bjarngard, "Head scatter off-axis for megavoltage x

rays," Med. Phys. **30**, 533–543 (2003).
[142]H. H. Liu, T. R. Mackie, and E. C. McCullough, "A dual source photon beam model used in convolution/superposition dose calculations for clinical megavoltage x-ray beams," Med. Phys. **24**, 1960–1974 (1997).
[143]M. B. Sharpe, D. A. Jaffray, J. J. Battista, and P. Munro, "Extrafocal radiation: A unified approach to the prediction of beam penumbra and output factors for megavoltage x-ray beams," Med. Phys. **22**, 2065–2074 (1995).
[144]H. H. Liu, T. R. Mackie, and E. C. McCullough, "Calculating output factors for photon beam radiotherapy using a convolution/superposition method based on a dual source photon beam model," Med. Phys. **24**, 1975–1985 (1997).
[145]M. R. Arnfield, J. V. Siebers, J. O. Kim, Q. Wu, P. J. Keall, and R. Mohan, "A method for determining multileaf collimator transmission and scatter for dynamic intensity modulated radiotherapy," Med. Phys. **27**, 2231–2241 (2000).
[146]H. H. Liu, T. R. Mackie, and E. C. McCullough, "Modeling photon output caused by backscattered radiation into the monitor chamber from collimator jaws using a Monte Carlo technique," Med. Phys. **27**, 737–744 (2000).
[147]S. B. Jiang, A. L. Boyer, and C. M. Ma, "Modeling the extrafocal radiation and monitor chamber backscatter for photon beam dose calculation," Med. Phys. **28**, 55–66 (2001).
[148]G. X. Ding, "Using Monte Carlo simulations to commission photon beam output factors—A feasibility study," Phys. Med. Biol. **48**, 3865–3874 (2003).
[149]B. Parker, A. S. Shiu, and H. H. Liu, "Small-field dosimetry with multiple detectors and Monte Carlo calculations," Med. Phys. **29**, 1372 (abstract) (2002).
[150]G. X. Ding, "Dose discrepancies between Monte Carlo calculations and measurements in the buildup region for a high-energy photon beam," Med. Phys. **29**, 2459–2463 (2002).
[151]G. X. Ding, C. Duzenli, and N. I. Kalach, "Are neutrons responsible for the dose discrepancies between Monte Carlo calculations and measurements in the build-up region for a high-energy photon beam?" Phys. Med. Biol. **47**, 3251–3261 (2002).
[152]W. Abdel-Rahman, J. P. Seuntjens, F. Verhaegen, F. Deblois, and E. B. Podgorsak, "Validation of Monte Carlo calculated surface doses for megavoltage photon beams," Med. Phys. **32**, 286–298 (2005).
[153]I. Kawrakow, "Efficient photon beam treatment head simulations," Radiother. Oncol. **81**, 82 (abstract) (2006).
[154]D. Sheikh-Bagheri, D. W. O. Rogers, C. K. Ross, and J. P. Seuntjens, "Comparison of measured and Monte Carlo calculated dose distributions from the NRC linac," Med. Phys. **27**, 2256–2266 (2000).
[155]I. Kawrakow, "On the effective point of measurement in megavoltage photon beams," Med. Phys. **33**, 1829–1839 (2006).
[156]O. Chibani and C. M. Ma, "On the discrepancies between Monte Carlo dose calculations and measurements for the 18 MV varian photon beam," Med. Phys. **34**, 1206–1216 (2007).
[157]W. Feller, *An Introduction to Probability Theory and Its Applications*, 3rd ed. (Wiley, New York, 1967), Vol. I.
[158]B. R. B. Walters, I. Kawrakow, and D. W. O. Rogers, "History by history statistical estimators in the BEAM code system," Med. Phys. **29**, 2745–2752 (2002).
[159]J. Sempau and A. F. Bielajew, "Towards the elimination of Monte Carlo statistical fluctuation from dose volume histograms for radiotherapy treatment planning," Phys. Med. Biol. **45**, 131–157 (2000).
[160]P. J. Keall, J. V. Siebers, R. Jeraj, and R. Mohan, "The effect of dose calculation uncertainty on the evaluation of radiotherapy plans," Med. Phys. **27**, 478–484 (2000).
[161]I. Kawrakow, "The effect of Monte Carlo statistical uncertainties on the evaluation of dose distributions in radiation treatment planning," Phys. Med. Biol. **49**, 1549–1556 (2004).
[162]I. J. Chetty, M. Rosu, M. L. Kessler, B. A. Fraass, R. K. Ten Haken, F. M. Kong, and D. L. McShan, "Reporting and analyzing statistical uncertainties in Monte Carlo-based treatment planning," Int. J. Radiat. Oncol. Biol. Phys. **65**, 1249–1259 (2006).
[163]J. V. Siebers, P. J. Keall, and I. Kawrakow, in *The Modern Technology of Radiation Oncology*, edited by J. Van Dyke (Medical Physics, Madison, WI, 2005), Vol. 2, pp. 91–130.
[164]F. M. Buffa and A. E. Nahum, "Monte Carlo dose calculations and radiobiological modelling: Analysis of the effect of the statistical noise of the dose distribution on the probability of tumour control," Phys. Med. Biol. **45**, 3009–3023 (2000).
[165]S. B. Jiang, T. Pawlicki, and C. M. Ma, "Removing the effect of statistical uncertainty on dose-volume histograms from Monte Carlo dose calculations," Phys. Med. Biol. **45**, 2151–2161 (2000).
[166]J. O. Deasy, "Denoising of electron beam Monte Carlo dose distributions using digital filtering techniques," Phys. Med. Biol. **45**, 1765–1779 (2000).
[167]I. Kawrakow, "On the de-noising of Monte Carlo calculated dose distributions," Phys. Med. Biol. **47**, 3087–3103 (2002).
[168]J. O. Deasy, M. V. Wickerhauser, and M. Picard, "Accelerating Monte Carlo simulations of radiation therapy dose distributions using wavelet threshold de-noising," Med. Phys. **29**, 2366–2373 (2002).
[169]S. J. Pollack and A. F. Bielajew, "Novel algorithms for smoothing global Monte Carlo noise," in Proceedings of the Current Topics in Monte Carlo Treatment Planning: Advanced Workshop, Montreal, CN, edited by F. Verhaegen and J. Seuntjens, 2004 (unpublished).
[170]B. Miao, R. Jeraj, S. Bao, and T. R. Mackie, "Adaptive anisotropic diffusion filtering of Monte Carlo dose distributions," Phys. Med. Biol. **48**, 2767–2781 (2003).
[171]M. Fippel and F. Nusslin, "Smoothing Monte Carlo calculated dose distributions by iterative reduction of noise," Phys. Med. Biol. **48**, 1289–1304 (2003).
[172]I. El Naqa *et al.*, "A comparison of Monte Carlo dose calculation denoising techniques," Phys. Med. Biol. **50**, 909–922 (2005).
[173]C. M. Ma, R. A. Price, Jr., J. S. Li, L. Chen, L. Wang, E. Fourkal, L. Qin, and J. Yang, "Monitor unit calculation for Monte Carlo treatment planning," Phys. Med. Biol. **49**, 1671–1687 (2004).
[174]F. C. du Plessis, C. A. Willemse, M. G. Lotter, and L. Goedhals, "The indirect use of CT numbers to establish material properties needed for Monte Carlo calculation of dose distributions in patients," Med. Phys. **25**, 1195–1201 (1998).
[175]C.-M. Ma and D. W. O. Rogers, "BEAMDP Users Manual," NRC Report PIRS-0509(D), 1995.
[176]J. V. Siebers, P. J. Keall, A. E. Nahum, and R. Mohan, "Converting absorbed dose to medium to absorbed dose to water for Monte Carlo based photon beam dose calculations," Phys. Med. Biol. **45**, 983–995 (2000).
[177]ICRU-Report No. 46: Photon, electron, proton and neutron interaction data for body tissues," in International Commission on Radiation Units and Measurements, 1992.
[178]F. Verhaegen and S. Devic, "Sensitivity study for CT image use in Monte Carlo treatment planning," Phys. Med. Biol. **50**, 937–946 (2005).
[179]M. Bazalova, L. Beaulieu, S. Palefsky, and F. Verhaegen, "Correction of CT artifacts and its influence on Monte Carlo dose calculations," Med. Phys. **34**, 2119–2132 (2007).
[180]C. Reft *et al.*, "Dosimetric considerations for patients with HIP prostheses undergoing pelvic irradiation. Report of the AAPM Radiation Therapy Committee Task Group 63," Med. Phys. **30**, 1162–1182 (2003).
[181]N. Dogan, J. V. Siebers, and P. J. Keall, "Clinical comparison of head and neck and prostate IMRT plans using absorbed dose to medium and absorbed dose to water," Phys. Med. Biol. **51**, 4967–4980 (2006).
[182]H. H. Liu, "D_m rather than D_w should be used in Monte Carlo treatment planning. For the proposition," Med. Phys. **29**, 922–923 (2002).
[183]M. Goitein, "The cell's-eye view: assessing dose in four dimensions," Int. J. Radiat. Oncol. Biol. Phys. **62**, 951–953 (2005).
[184]M. Fippel and F. Nusslin, "Comments on 'Converting absorbed dose to medium to absorbed dose to water for Monte Carlo based photon beam dose calculations'," Phys. Med. Biol. **45**, L17–L19 (2000).
[185]J. Siebers, B. Libby, and R. Mohan, "Trust, but verify: Comparison of MCNP and BEAM Monte Carlo codes for generation of phase space distributions for a Varian 2100C," Med. Phys. **25**, A143 (abstract) (1998).
[186]R. Mohan, M. Arnfield, S. Tong, Q. Wu, and J. Siebers, "The impact of fluctuations in intensity patterns on the number of monitor units and the quality and accuracy of intensity modulated radiotherapy," Med. Phys. **27**, 1226–1237 (2000).
[187]J. O. Kim, J. V. Siebers, P. J. Keall, M. R. Arnfield, and R. Mohan, "A Monte Carlo study of radiation transport through multileaf collimators," Med. Phys. **28**, 2497–2506 (2001).
[188]P. J. Keall, J. V. Siebers, M. Arnfield, J. O. Kim, and R. Mohan, "Monte Carlo dose calculations for dynamic IMRT treatments," Phys. Med. Biol. **46**, 929–941 (2001).
[189]J. V. Siebers, M. Lauterbach, P. J. Keall, and R. Mohan, "Incorporating multi-leaf collimator leaf sequencing into iterative IMRT optimization,"

Med. Phys. **29**, 952–959 (2002).
[190]L. Wang, E. Yorke, and C. S. Chui, "Monte Carlo evaluation of 6 MV intensity modulated radiotherapy plans for head and neck and lung treatments," Med. Phys. **29**, 2705–2717 (2002).
[191]C.-M. Ma *et al.*, "Monte Carlo verification of IMRT dose distributions from a commercial treatment planning optimization system," Phys. Med. Biol. **45**, 2483–2495 (2000).
[192]T. Pawlicki and C. M. Ma, "Monte Carlo simulation for MLC-based intensity-modulated radiotherapy," Med. Dosim. **26**, 157–168 (2001).
[193]P. Francescon, S. Cora, and P. Chiovati, "Dose verification of an IMRT treatment planning system with the BEAM EGS4-based Monte Carlo code," Med. Phys. **30**, 144–157 (2003).
[194]M. Rincon *et al.*, "Monte Carlo conformal treatment planning as an independent assessment," in *Advanced Monte Carlo for Radiation Physics: Proceedings of the Monte Carlo 2000 Meeting, Lisbon*, edited by A. Kling *et al.* (Springer-Verlag, Berlin, 2001), pp. 565–570.
[195]R. Jeraj, P. J. Keall, and J. V. Siebers, "The effect of dose calculation accuracy on inverse treatment planning," Phys. Med. Biol. **47**, 391–407 (2002).
[196]N. Reynaert *et al.*, "The importance of accurate linear accelerator head modelling for IMRT Monte Carlo calculations," Phys. Med. Biol. **50**, 831–846 (2005).
[197]W. Laub, M. Alber, M. Birkner, and F. Nusslin, "Monte Carlo dose computation for IMRT optimization," Phys. Med. Biol. **45**, 1741–1754 (2000).
[198]N. Dogan, J. V. Siebers, P. J. Keall, F. Lerma, Y. Wu, M. Fatyga, J. F. Williamson, and R. K. Schmidt-Ullrich, "Improving IMRT dose accuracy via deliverable Monte Carlo optimization for the treatment of head and neck cancers," Med. Phys. **33**, 4033–4055 (2006).
[199]J. V. Siebers, M. Lauterbach, S. Tong, Q. Wu, and R. Mohan, "Reducing dose calculation time for accurate iterative IMRT planning," Med. Phys. **29**, 231–237 (2002).
[200]A. M. Bergman, K. Bush, M. P. Milette, I. A. Popescu, K. Otto, and C. Duzenli, "Direct aperture optimization for IMRT using Monte Carlo generated beamlets," Med. Phys. **33**, 3666–3679 (2006).
[201]B. De Smedt, B. Vanderstraeten, N. Reynaert, W. De Neve, and H. Thierens, "Investigation of geometrical and scoring grid resolution for Monte Carlo dose calculations for IMRT," Phys. Med. Biol. **50**, 4005–4019 (2005).
[202]J. F. Dempsey, H. E. Romeijn, J. G. Li, D. A. Low, and J. R. Palta, "A Fourier analysis of the dose grid resolution required for accurate IMRT fluence map optimization," Med. Phys. **32**, 380–388 (2005).
[203]J. H. Hubbell, "Review of photon interaction cross section data in the medical and biological context," Phys. Med. Biol. **44**, R1–R22 (1999).
[204]D. V. Rao, S. M. Seltzer, and P. M. Bergstrom, Jr., "Compton scattering cross-sections for individual subshells for a few elements of biological interest in the energy region 5 keV–10 MeV," Radiat. Phys. Chem. **70**, 479–489 (2004).
[205]S. M. Seltzer, in *Monte Carlo Transport of Electrons and Photons*, edited by W. R. Nelson, T. M. Jenkins, A. Rindi, A. E. Nahum, and D. W. O. Rogers (Plenum, New York, 1988), pp. 81–114.
[206]B. A. Fraass, J. Smathers, and J. Deye, "Summary and recommendations of a National Cancer Institute workshop on issues limiting the clinical use of Monte Carlo dose calculation algorithms for megavoltage external beam radiation therapy," Med. Phys. **30**, 3206–3216 (2003).
[207]E. R. Epp, A. L. Boyer, and K. P. Doppke, "Underdosing of lesions resulting from lack of electronic equilibrium in upper respiratory air cavities irradiated by 10 MV x-ray beams," Int. J. Radiat. Oncol. Biol. Phys. **2**, 613–619 (1977).
[208]M. A. Hunt, G. E. Desobry, B. Fowble, and L. R. Coia, "Effect of low-density lateral interfaces on soft-tissue doses," Int. J. Radiat. Oncol. Biol. Phys. **37**, 475–482 (1997).
[209]E. E. Klein, L. M. Chin, R. K. Rice, and B. J. Mijnheer, "The influence of air cavities on interface doses for photon beams," Int. J. Radiat. Oncol. Biol. Phys. **27**, 419–427 (1993).
[210]T. R. Mackie, J. W. Scrimger, and J. J. Battista, "A convolution method of calculating dose for 15-MV x rays," Med. Phys. **12**, 188–196 (1985).
[211]R. Mohan, C. Chui, and L. Lidofsky, "Differential pencil beam dose computation model for photons," Med. Phys. **13**, 64–73 (1986).
[212]C. X. Yu, J. W. Wong, and J. A. Purdy, "Photon dose perturbations due to small inhomogeneities," Med. Phys. **14**, 78–83 (1987).
[213]M. R. Arnfield, C. H. Siantar, J. Siebers, P. Garmon, L. Cox, and R. Mohan, "The impact of electron transport on the accuracy of computed dose," Med. Phys. **27**, 1266–1274 (2000).
[214]F. C. du Plessis, C. A. Willemse, M. G. Lotter, and L. Goedhals, "Comparison of the Batho, ETAR and Monte Carlo dose calculation methods in CT based patient models," Med. Phys. **28**, 582–589 (2001).
[215]A. O. Jones and I. J. Das, "Comparison of inhomogeneity correction algorithms in small photon fields," Med. Phys. **32**, 766–776 (2005).
[216]T. Knoos, E. Wieslander, L. Cozzi, C. Brink, A. Fogliata, D. Albers, H. Nystrom, and S. Lassen, "Comparison of dose calculation algorithms for treatment planning in external photon beam therapy for clinical situations," Phys. Med. Biol. **51**, 5785–5807 (2006).
[217]M. Miften, M. Wiesmeyer, A. Kapur, and C. M. Ma, "Comparison of RTP dose distributions in heterogeneous phantoms with the BEAM Monte Carlo simulation system," J. Appl. Clin. Med. Phys. **2**, 21–31 (2001).
[218]R. Mohan, "Why Monte Carlo?," in *Proceedings of the 12th ICCR*, edited by D. Leavitt (Medical Physics, Salt Lake City, UT, 1997), pp. 16–18.
[219]S. N. Rustgi, A. K. Rustgi, S. B. Jiang, and K. M. Ayyangar, "Dose perturbation caused by high-density inhomogeneities in small beams in stereotactic radiosurgery," Phys. Med. Biol. **43**, 3509–3518 (1998).
[220]P. Carrasco *et al.*, "Comparison of dose calculation algorithms in phantoms with lung equivalent heterogeneities under conditions of lateral electronic disequilibrium," Med. Phys. **31**, 2899–2911 (2004).
[221]J. Coleman, C. Joy, J. E. Park, P. Villarreal-Barajas, P. L. Petti, and B. Faddegon, "A comparison of Monte Carlo and Fermi–Eyges–Hogstrom estimates of heart and lung dose from breast electron boost treatment," Int. J. Radiat. Oncol. Biol. Phys. **61**, 621–628 (2005).
[222]T. Krieger and O. A. Sauer, "Monte Carlo-versus pencil-beam-/collapsed-cone dose calculation in a heterogeneous multi-layer phantom," Phys. Med. Biol. **50**, 859–868 (2005).
[223]W. U. Laub, A. Bakai, and F. Nusslin, "Intensity modulated irradiation of a thorax phantom: Comparisons between measurements, Monte Carlo calculations and pencil beam calculations," Phys. Med. Biol. **46**, 1695–1706 (2001).
[224]E. Spezi, D. G. Lewis, and C. W. Smith, "Monte Carlo simulation and dosimetric verification of radiotherapy beam modifiers," Phys. Med. Biol. **46**, 3007–3029 (2001).
[225]L. Wang, M. Lovelock, and C. S. Chui, "Experimental verification of a CT-based Monte Carlo dose-calculation method in heterogeneous phantoms," Med. Phys. **26**, 2626–2634 (1999).
[226]K. De Vlamynck, H. Palmans, F. Verhaegen, C. De Wagter, W. De Neve, and H. Thierens, "Dose measurements compared with Monte Carlo simulations of narrow 6 MV multileaf collimator shaped photon beams," Med. Phys. **26**, 1874–1882 (1999).
[227]G. A. Ezzell *et al.*, "Guidance document on delivery, treatment planning, and clinical implementation of IMRT: Report of the IMRT Subcommittee of the AAPM Radiation Therapy Committee," Med. Phys. **30**, 2089–2115 (2003).
[228]C. G. Orton, P. M. Mondalek, J. T. Spicka, D. S. Herron, and L. I. Andres, "Benchmark measurements for lung dose corrections for x-ray beams," Int. J. Radiat. Oncol. Biol. Phys. **10**, 2191–2199 (1984).
[229]R. K. Rice, B. J. Mijnheer, and L. M. Chin, "Benchmark measurements for lung dose corrections for x-ray-beams," Int. J. Radiat. Oncol. Biol. Phys. **15**, 399–409 (1988).
[230]P. M. Charland, I. J. Chetty, S. Yokoyama, and B. A. Fraass, "Dosimetric comparison of extended dose range film with ionization measurements in water and lung equivalent heterogeneous media exposed to megavoltage photons," J. Appl. Clin. Med. Phys. **4**, 25–39 (2003).
[231]P. Dunscombe, P. McGhee, and E. Lederer, "Anthropomorphic phantom measurements for the validation of a treatment planning system," Phys. Med. Biol. **41**, 399–411 (1996).
[232]S. Yokoyama, P. L. Roberson, D. L. Litzenberg, J. M. Moran, and B. A. Fraass, "Surface buildup dose dependence on photon field delivery technique for IMRT," J. Appl. Clin. Med. Phys. **5**, 71–81 (2004).
[233]T. Kron, A. Elliot, T. Wong, G. Showell, B. Clubb, and P. Metcalfe, "X-ray surface dose measurements using TLD extrapolation," Med. Phys. **20**, 703–711 (1993).
[234]B. E. Bjarngard, P. Vadash, and T. Zhu, "Doses near the surface in high-energy x-ray beams," Med. Phys. **22**, 465–468 (1995).
[235]A. S. Shiu *et al.*, "Verification data for electron beam dose algorithms," Med. Phys. **19**, 623–636 (1992).
[236]R. A. Boyd, K. R. Hogstrom, J. A. Antolak, and A. S. Shiu, "A measured data set for evaluating electron-beam dose algorithms," Med. Phys. **28**, 950–958 (2001).
[237]K. De Jaeger, M. S. Hoogeman, M. Engelsman, Y. Seppenwoolde, E. M.

F. Damen, B. J. Mijnheer, L. J. Boersma, and J. V. Lebesque, "Incorporating an improved dose-calculation algorithm in conformal radiotherapy of lung cancer: Re-evaluation of dose in normal lung tissue," Radiother. Oncol. **69**, 1–10 (2003).

[238]I. J. Chetty, M. Rosu, F.-M. Kong, C. Lopez, D. S. Tatro, D. L. McShan, B. A. Fraass, and R. K. Ten Haken, "On the correlation of dose-volume-response using Monte Carlo dose calculation in conformal radiation therapy of lung cancer," in *Proceedings of the 14th ICCR*, edited by B.Y. Yi, S. D. Ahn, E. K. Choi, and S. W. Ha (Jeong, Seoul, Korea, 2004), pp. 457–460.

[239]P. E. Lindsay, I. El Naqa, A. Hope, M. Vicic, J. Cui, J. Bradley, and J. O. Deasy, "Retrospective Monte Carlo dose calculations with limited beam weight information," Med. Phys. **34**, 334–346 (2007).

[240]P. J. Keall, J. Siebers, and R. Mohan, "The impact of Monte Carlo dose calculations on treatment outcomes," in *Proceedings of the 13th ICCR*, edited by T. Bortfeld and W. Schlegel (Springer-Verlag, Heidelberg, 2000), pp. 425–427.

[241]N. Reynaert *et al.*, "Monte Carlo treatment planning for photon and electron beams," Radiat. Phys. Chem. **76**, 643–686 (2007).

[242]A. Fogliata, E. Vanetti, D. Albers, C. Brink, A. Clivio, T. Knoos, G. Nicolini, and L. Cozzi, "On the dosimetric behaviour of photon dose calculation algorithms in the presence of simple geometric heterogeneities: Comparison with Monte Carlo calculations," Phys. Med. Biol. **52**, 1363–1385 (2007).

[243]T. Knoos, A. Ahnesjo, P. Nilsson, and L. Weber, "Limitations of a pencil beam approach to photon dose calculations in lung tissue," Phys. Med. Biol. **40**, 1411–1420 (1995).

[244]P. N. McDermott, T. He, and A. DeYoung, "Dose calculation accuracy of lung planning with a commercial IMRT treatment planning system," J. Appl. Clin. Med. Phys. **4**, 341–351 (2003).

[245]M. F. Tsiakalos, K. Theodorou, C. Kappas, S. Zefkili, and J. C. Rosenwold, "Analysis of the penumbra enlargement in lung versus the quality index of photon beams: A methodology to check the dose calculation algorithm," Med. Phys. **31**, 943–949 (2004).

[246]E. D. Yorke, L. Wang, K. E. Rosenzweig, D. Mah, J. B. Paoli, and C. S. Chui, "Evaluation of deep inspiration breath-hold lung treatment plans with Monte Carlo dose calculation," Int. J. Radiat. Oncol. Biol. Phys. **53**, 1058–1070 (2002).

[247]R. Timmerman, L. Papiez, R. McGarry, L. Likes, C. DesRosiers, S. Frost, and M. Williams, "Extracranial stereotactic radioablation: results of a phase I study in medically inoperable stage I non-small cell lung cancer," Chest **124**, 1946–1955 (2003).

[248]L. Wang, E. Yorke, and C. S. Chui, "Monte Carlo evaluation of tissue inhomogeneity effects in the treatment of the head and neck," Int. J. Radiat. Oncol. Biol. Phys. **50**, 1339–1349 (2001).

[249]C. Boudreau, E. Heath, J. Seuntjens, O. Ballivy, and W. Parker, "IMRT head and neck treatment planning with a commercially available Monte Carlo based planning system," Phys. Med. Biol. **50**, 879–890 (2005).

[250]J. Seco, E. Adams, M. Bidmead, M. Partridge, and F. Verhaegen, "Head-and-neck IMRT treatments assessed with a Monte Carlo dose calculation engine," Phys. Med. Biol. **50**, 817–830 (2005).

[251]C. Martens, N. Reynaert, C. De Wagter, P. Nilsson, M. Coghe, H. Palmans, H. Thierens, and W. De Neve, "Underdosage of the upper-airway mucosa for small fields as used in intensity-modulated radiation therapy: A comparison between radiochromic film measurements, Monte Carlo simulations, and collapsed cone convolution calculations," Med. Phys. **29**, 1528–1535 (2002).

[252]F. Verhaegen, I. J. Das, and H. Palmans, "Monte Carlo dosimetry study of a 6 MV stereotactic radiosurgery unit," Phys. Med. Biol. **43**, 2755–2768 (1998).

[253]F. Sánchez-Doblado *et al.*, "Ionization chamber dosimetry of small photon fields: A Monte Carlo study on stopping-power ratios for radiosurgery and IMRT beams," Phys. Med. Biol. **48**, 2081–2099 (2003).

[254]A. Chaves, M. C. Lopes, C. C. Alves, C. Oliveira, L. Peralta, P. Rodrigues, and A. Trindade, "A Monte Carlo multiple source model applied to radiosurgery narrow photon beams," Med. Phys. **31**, 2192–2204 (2004).

[255]J. S. Li *et al.*, "Clinical implementation of intensity-modulated tangential beam irradiation for breast cancer," Med. Phys. **31**, 1023–1031 (2004).

[256]J. J. DeMarco, I. J. Chetty, and T. D. Solberg, "A Monte Carlo tutorial and the application for radiotherapy treatment planning," Med. Dosim. **27**, 43–50 (2002).

[257]A. Leal, F. Sanchez-Doblado, R. Arrans, J. Rosello, E. C. Pavon, and J. I. Lagares, "Routine IMRT verification by means of an automated Monte Carlo simulation system," Int. J. Radiat. Oncol. Biol. Phys. **56**, 58–68 (2003).

[258]G. X. Ding, D. M. Duggan, C. W. Coffey, P. Shokrani, and J. E. Cygler, "First macro Monte Carlo based commercial dose calculation module for electron beam treatment planning—New issues for clinical consideration," Phys. Med. Biol. **51**, 2781–2799 (2006).

[259]D. W. O. Rogers, A. F. Bielajew, and A. E. Nahum, "Monte Carlo calculations of electron beams in standard dose planning geometries," in *Proceedings of the 8th ICCR* (IEEE, New York, 1984), pp. 140–144.

[260]C. Scherf, J. Scherer, and L. Bogner, "Verification and application of the Voxel-based Monte Carlo (VMC++) electron dose module of oncentratrade mark MasterPlan," Strahlenther. Onkol. **183**, 81–88 (2007).

[261]K. R. Shortt, C. K. Ross, A. F. Bielajew, and D. W. O. Rogers, "Electron beam dose distributions near standard inhomogeneities," Phys. Med. Biol. **31**, 235–249 (1986).

[262]E. Wieslander and T. Knoos, "A virtual-accelerator-based verification of a Monte Carlo dose calculation algorithm for electron beam treatment planning in clinical situations," Radiother. Oncol. **82**, 208–217 (2007).

[263]J. Cygler, J. J. Battista, J. W. Scrimger, E. Mah, and J. Antolak, "Electron dose distributions in experimental phantoms: a comparison with 2D pencil beam calculations," Phys. Med. Biol. **32**, 1073–1086 (1987).

[264]J. A. Hayman *et al.*, "Dose escalation in non-small-cell lung cancer using three-dimensional conformal radiation therapy: Update of a phase I trial," J. Clin. Oncol. **19**, 127–136 (2001).

[265]I. J. Chetty, M. Rosu, D. L. McShan, B. A. Fraass, and R. K. Ten Haken, "The influence of beam model differences in the comparison of dose calculation algorithms for lung cancer treatment planning," Phys. Med. Biol. **50**, 801–815 (2005).

[266]F. M. Kong, R. K. Ten Haken, M. J. Schipper, M. A. Sullivan, M. Chen, C. Lopez, G. P. Kalemkerian, and J. A. Hayman, "High-dose radiation improved local tumor control and overall survival in patients with inoperable/unresectable non-small-cell lung cancer: Long-term results of a radiation dose escalation study," Int. J. Radiat. Oncol. Biol. Phys. **63**, 324–333 (2005).

[267]I. J. Chetty, "Monte Carlo treatment planning: The influence of 'variance reduction' techniques (ECUT, PCUT, ESTEP) on the accuracy and speed of dose calculations," Med. Phys. **32**, 2018 (abstract) (2005).

[268]www.irs.inms.nrc.ca/inms/irs/papers/iccr00/iccr00.html.

[269]E. Heath, McGill University (personal communication).

[270]E. Poon and F. Verhaegen, McGill University (personal communication).

A Fourier analysis of the dose grid resolution required for accurate IMRT fluence map optimization

James F. Dempsey[a)]
Department of Radiation Oncology, University of Florida College of Medicine, Gainesville, Florida 32610-0385

H. Edwin Romeijn
Department of Industrial and Systems Engineering, University of Florida, Gainesville, Florida 32611–6595

Jonathan G. Li
Department of Radiation Oncology, University of Florida College of Medicine, Gainesville, Florida 32610-0385

Daniel A. Low
Department of Radiation Oncology, Washington University School of Medicine, St. Louis, Missouri 63110–1032

Jatinder R. Palta
Department of Radiation Oncology, University of Florida College of Medicine, Gainesville, Florida 32610-0385

(Received 10 July 2004; revised 8 November 2004; accepted for publication 9 November 2004; published 14 January 2005)

We present a theoretical and empirical analysis of the errors associated with the spatial discretization of the dose grid employed in optimized intensity modulated radiation therapy (IMRT) treatment plans. An information theory based Fourier analysis of the accuracy of discrete representations of three-dimensional dose distributions is presented. When applied to beamlet-based IMRT dose distributions, the theory produces analytic integrals that can bound worst case aliasing errors that can occur regardless of the location and orientation of the dose grid. The predictions of this theory are compared to empirical results obtained by solving a linear-programming based fluence-map optimization model to global optimality. A reasonable agreement between worst case estimates and the empirical results is attributed to the fact that the optimization takes advantage of aliasing to produce an optimal plan. We predicted and empirically demonstrated that an isotropic dose grid with <2.5 mm spacing is sufficient to prevent dose errors larger than a percent. However, we noted that in practice this resolution is mostly needed in high-dose target regions. Finally, a multiresolution 2–4–6 mm spacing model was developed and empirically tested where these spacings were applied to targets, structures, and tissue, respectively. © *2005 American Association of Physicists in Medicine.* [DOI: 10.1118/1.1843354]

Key words: IMRT, treatment plan evaluation, optimization

I. INTRODUCTION

When implementing megavoltage (MV) photon-beam intensity modulated radiation therapy (IMRT) achieved through multileaf collimator (MLC) delivery, tumorcidal high-dose regions and tissue sparing low-dose regions are constructed via the superposition of many small MLC-shaped fields. For conventional conformal therapy, the characteristics of the penumbra of a radiation field are of little consequence as uniform or one-dimensionally modulated (wedged) fields are designed to wholly cover targets with an adequate margin to allow for penumbra fall off. In fact, the matching of fields over targets in conventional conformal therapy is avoided if possible or performed with great consideration for placement, utilizing matchline-feathering techniques to minimize hot and cold spots. However, IMRT represents a significant departure from previous practice, where the ability to shape dose distributions relies heavily on the use and characterization of the penumbra of many irregular MLC fields, which are purposely and necessarily matched over the target. This shift in radiotherapy delivery practice demands an appropriate adjustment of the standards of treatment planning algorithms to meet the challenge. In the development of three-dimensional conformal radiation therapy over a decade ago, the impact of discretization of the computed dose distributions on the dosimetric accuracy for treatment planning was studied by the analysis of empirical analytic models of dose distributions.[1–7] Given the greater relevance of the penumbra with its steep gradients in MLC based IMRT, we have revisited this issue.

Recently, Bortfeld *et al.*,[8] have presented a very elegant and general Fourier analysis of the discrete MLC size to employ in the design of IMRT MLC systems by considering their resultant dose distributions. In this study, we develop a general analysis to determine the worst case discretization errors possible when employing discretized dose distribu-

tions for IMRT fluence map optimization (FMO). Worst case error analysis proves very useful as it puts an absolute limit on errors and greatly simplifies the computational complexity of the error analysis. Similar to the work of Bortfeld *et al.*,[8] we employ an information theory[9] and Fourier analysis[10] based approach.

II. MATERIALS AND METHODS

A. Information theory and discretization

A general theory of the accuracy of discretized approximation of a continuous three-dimensional (3D) function, $f(\mathbf{r})$, where $\mathbf{r}=(x,y,z)$ represents a Cartesian spatial coordinate in 3D space, can be based on the well known Fourier transform, $\Im$, of that function, $\tilde{f}(\boldsymbol{\varpi})$:

$$\tilde{f}(\boldsymbol{\varpi}) = \Im[f(\mathbf{r})] = \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} f(\mathbf{r}) e^{2\pi i \boldsymbol{\varpi}\cdot\mathbf{r}} d^3\mathbf{r}, \tag{1}$$

where $\boldsymbol{\varpi}$ represents a Cartesian frequency coordinate and the corresponding inverse Fourier transform, $\Im^{-1}$, of $f(\mathbf{r})$ is given by

$$f(\mathbf{r}) = \Im^{-1}[\tilde{f}(\boldsymbol{\varpi})] = \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} \tilde{f}(\boldsymbol{\varpi}) e^{-2\pi i \boldsymbol{\varpi}\cdot\mathbf{r}} d^3\boldsymbol{\varpi}. \tag{2}$$

From the Nyquist–Shannon theorem (see, e.g., Chap. 10 of Bracewell[11]) we know that the continuous function f can be exactly described by a discretized version of the function, say $\check{f}_h$, defined on a regular isotropic grid with grid-spacing $h=1/2\Omega$, if and only if

$$\begin{aligned} f(\mathbf{r}) &= \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} \tilde{f}(\boldsymbol{\varpi}) e^{-2\pi i \boldsymbol{\varpi}\cdot\mathbf{r}} d^3\boldsymbol{\varpi} \\ &= \int_{-\Omega}^{+\Omega}\int_{-\Omega}^{+\Omega}\int_{-\Omega}^{+\Omega} \tilde{f}(\boldsymbol{\varpi}) e^{-2\pi i \boldsymbol{\varpi}\cdot\mathbf{r}} d^3\boldsymbol{\varpi}. \end{aligned} \tag{3}$$

In other words, if there is some (typically large) frequency, Ω, which *band limits* f, then any $\check{f}_h$ evaluated on an isotropic grid with a spacing of $h \leq 1/2\Omega$ will contain all of the information in f. This means that $\check{f}_h$ will exactly represent f and Ω is typically called the Nyquist frequency of f. Use of the method of Fourier interpolation applied to the function $\check{f}_h$ will allow for an exact recovery of f at points that are not on the specified grid. However, it may safely be assumed that most other reasonable models of interpolation will yield accurate results from Nyquist–Shannon limit sampled functions.

We now note that different types of discretization are often employed in approximating a hypothetical continuous 3D dose distribution function,[12] d. (We also note that it is well known that absorbed ionizing radiation dose distributions are not even theoretically continuous and well behaved functions at a microscopic scale due to discrete particle interactions giving rise to the absorbed ionizing dose. However, there should be a well behaved function of the expectation of the mean dose that is taken as a surrogate for the actual stochastic dose at some mesoscopic scale that is completely adequate for the purposes of radiation oncology. Such a coarse graining of the problem is justifiable on the order of the size of a cell as this is the intended target of radiotherapy.) To deal with this, we present a formal description for the two mechanisms of discretization most commonly encountered in dosimetric modeling for radiation therapy. We can model the process of discretizing d by considering mathematical functions and transforms that mimic the physical operations of discretization. Typically, there are two types of discretization that need to be considered: (1) discrete sampling and (2) volumetric averaging. Examples of each type of discretization would include (1) estimation of an analytic dose distribution model on a grid of points to approximate this dose distribution and (2) estimation of a stochastic dose distribution model as the average value of absorbed dose inside voxels using Monte Carlo simulation, respectively. We seek to answer the questions: "How do these different types of discretization relate to each other and influence the estimated dose distribution?"

To deal with the case of discrete sampling at points we define a Cartesian 3D sampling function on a grid with isotropic unit spacing

$$\mathrm{III}(\mathbf{r}) = \sum_{i=-\infty}^{\infty}\sum_{j=-\infty}^{\infty}\sum_{k=-\infty}^{\infty} \delta(x-i, y-j, z-k), \tag{4}$$

where i, j, and k are integers and δ is the Dirac delta function

$$\delta(\mathbf{r}) = \lim_{\epsilon\to 0} \frac{1}{2\sqrt{\pi\epsilon}} e^{-|\mathbf{r}|^2/4\epsilon}. \tag{5}$$

The function III is also known as the *Shah* function and it is a well-known result that is its own Fourier transform and has the following scaling property:[10]

$$\mathrm{III}_h(\mathbf{r}) = |h^3| \sum_{i=-\infty}^{\infty}\sum_{j=-\infty}^{\infty}\sum_{k=-\infty}^{\infty} \delta(x-hi, y-hj, z-hk), \tag{6}$$

where h is an isotropic grid spacing. We now observe that the operation of discrete sampling on an isotropic grid with spacing h to produce a discretized approximation of a continuous 3D dose distribution function, $\check{d}_h$, consists of taking the product of the continuous 3D dose distribution function d and III_h:

$$\check{d}_h(\mathbf{r}) = \mathrm{III}_h(\mathbf{r}) \times d(\mathbf{r}). \tag{7}$$

To deal with the case of volumetric averaging we define a unit Cartesian 3D voxel function

$$\begin{aligned} {}^3\Pi(\mathbf{r}) = &[H(x+\tfrac{1}{2}) - H(x-\tfrac{1}{2})][H(y+\tfrac{1}{2}) - H(y-\tfrac{1}{2})] \\ &\times [H(z+\tfrac{1}{2}) - H(z-\tfrac{1}{2})], \end{aligned} \tag{8}$$

where H is the heaviside unit step function:

$$H(r)=\begin{cases}0 & r<0\\ \frac{1}{2} & r=0\\ 1 & r>0.\end{cases} \tag{9}$$

Analogously, we can define the function ${}^3\Pi_h$ for the case of voxels of isotropic dimension and grid spacing h. We note that ${}^3\Pi$ has the following scaling property

$${}^3\Pi_h(\mathbf{r})=|h|^{-3}\,{}^3\Pi\left(\frac{\mathbf{r}}{h}\right) \tag{10}$$

and that the Fourier transform of ${}^3\Pi_h$ is well known and given by

$${}^3\tilde{\Pi}_h(\kappa,\lambda,\mu)=\mathrm{sinc}(h\kappa)\mathrm{sinc}(h\lambda)\mathrm{sinc}(h\mu), \tag{11}$$

where the sinc or *sine cardinal* function is defined as

$$\mathrm{sinc}(r)=\frac{\sin(\pi r)}{\pi r} \tag{12}$$

and κ, λ, and μ are the Cartesian frequency coordinates corresponding to x, y, and z. The voxel function ${}^3\Pi_h$ represents a cubic region of volume h^3. By convolving d with ${}^3\Pi_h$ we obtain a volume averaged dose distribution

$$d_h^{\mathrm{VA}}(\mathbf{r})=\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} d(\mathbf{r}-\mathbf{r}')^3\Pi_h(\mathbf{r}')d^3\mathbf{r}'. \tag{13}$$

Now the operation of discrete sampling can be applied to produce a discretized approximation $\check{d}_h^{\mathrm{VA}}$ of the volume averaged continuous function d_h^{VA} by

$$\check{d}_h^{\mathrm{VA}}(\mathbf{r})=\mathrm{III}_h(\mathbf{r})\times d_h^{\mathrm{VA}}(\mathbf{r}), \tag{14}$$

where we have assumed that the sampling grid and voxel dimension are both equal to h. Finally, we note that one can obtain d from $\check{d}_h^{\mathrm{VA}}$ by deconvolution with ${}^3\Pi_h$ as long as the isotropic grid spacing is Nyquist–Shannon limited.

Acknowledging that one cannot always expect to use a grid spacing small enough to truly band limit d and that one would always like to compute and store $\check{d}_h$ (or $\check{d}_h^{\mathrm{VA}}$) as efficiently as possible, we then seek to evaluate the discrepancy between the dose distribution d and its discretizations $\check{d}_h$ and $\check{d}_h^{\mathrm{VA}}$. To this end, we first consider the following decomposition of the inverse Fourier transform [from Eq. (2)] applied to the dose distribution functions d and d_h^{VA} in a Cartesian coordinate frame

$$\begin{aligned} d(\mathbf{r})&=\sum_{i=1}^{3}\sum_{j=1}^{3}\sum_{k=1}^{3}\int_{A_i}\int_{A_j}\int_{A_k}\tilde{d}(\boldsymbol{\varpi})e^{-2\pi i\boldsymbol{\varpi}\cdot\mathbf{r}}d^3\boldsymbol{\varpi},\\ d_h^{\mathrm{VA}}(\mathbf{r})&=\sum_{i=1}^{3}\sum_{j=1}^{3}\sum_{k=1}^{3}\int_{A_i}\int_{A_j}\int_{A_k}\tilde{d}(\boldsymbol{\varpi})\,{}^3\tilde{\Pi}_h(\boldsymbol{\varpi})e^{-2\pi i\boldsymbol{\varpi}\cdot\mathbf{r}}d^3\boldsymbol{\varpi}, \end{aligned} \tag{15}$$

where $A_1=(-\infty,-1/2h)$, $A_2=[-1/2h,1/2h]$, and $A_3=(1/2h,\infty)$. Noting that

$$\begin{aligned} \check{d}_h(\mathbf{r})&=\int_{A_2}\int_{A_2}\int_{A_2}\tilde{d}(\boldsymbol{\varpi})e^{-2\pi i\boldsymbol{\varpi}\cdot\mathbf{r}}d^3\boldsymbol{\varpi},\\ \check{d}_h^{\mathrm{VA}}(\mathbf{r})&=\int_{A_2}\int_{A_2}\int_{A_2}\tilde{d}(\boldsymbol{\varpi})\,{}^3\tilde{\Pi}_h(\boldsymbol{\varpi})e^{-2\pi i\boldsymbol{\varpi}\cdot\mathbf{r}}d^3\boldsymbol{\varpi} \end{aligned} \tag{16}$$

we can now study the discrepancy functions $\Delta_h=d-\check{d}_h$:

$$\begin{aligned} \Delta_h(\mathbf{r})=&\left(\sum_{i=1}^{3}\sum_{j=1}^{3}\sum_{k=1}^{3}\int_{A_i}\int_{A_j}\int_{A_k}\tilde{d}(\boldsymbol{\varpi})e^{-2\pi i\mathbf{r}\cdot\boldsymbol{\varpi}}d^3\boldsymbol{\varpi}\right)\\ &-\int_{A_2}\int_{A_2}\int_{A_2}\tilde{d}(\boldsymbol{\varpi})e^{-2\pi i\mathbf{r}\cdot\boldsymbol{\varpi}}d^3\boldsymbol{\varpi} \end{aligned} \tag{17}$$

and $\Delta_h^{\mathrm{VA}}=d_h^{\mathrm{VA}}-\check{d}_h^{\mathrm{VA}}$:

$$\begin{aligned} \Delta_h^{\mathrm{VA}}(\mathbf{r})=&\left(\sum_{i=1}^{3}\sum_{j=1}^{3}\sum_{k=1}^{3}\int_{A_i}\int_{A_j}\int_{A_k}\tilde{d}(\boldsymbol{\varpi})\,{}^3\tilde{\Pi}_h(\boldsymbol{\varpi})e^{-2\pi i\mathbf{r}\cdot\boldsymbol{\varpi}}d^3\boldsymbol{\varpi}\right)\\ &-\int_{A_2}\int_{A_2}\int_{A_2}\tilde{d}(\boldsymbol{\varpi})\,{}^3\tilde{\Pi}_h(\boldsymbol{\varpi})e^{-2\pi i\mathbf{r}\cdot\boldsymbol{\varpi}}d^3\boldsymbol{\varpi}. \end{aligned} \tag{18}$$

Note that the last terms in Eqs. (17) and (18) remove the discrete parts of the function up to a frequency of $1/2h$.

It is fairly obvious that the discrepancies found between a continuous function and a discrete approximation to that function should depend on the alignment of the discretization grid with the continuous function. In particular, consider the impact of a spatial offset of a function f on its Fourier transform $\tilde{f}$:

$$\begin{aligned} &\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty}f(\mathbf{r}-\mathbf{a})e^{-2\pi i\boldsymbol{\varpi}\cdot\mathbf{r}}d^3\mathbf{r}\\ &=\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty}f(\mathbf{r}-\mathbf{a})e^{-2\pi i\boldsymbol{\varpi}\cdot(\mathbf{r}-\mathbf{a})}e^{-2\pi i\boldsymbol{\varpi}\cdot\mathbf{a}}d^3(\mathbf{r}-\mathbf{a})\\ &=e^{-2\pi i\boldsymbol{\varpi}\cdot\mathbf{a}}\tilde{f}(\boldsymbol{\varpi}), \end{aligned} \tag{19}$$

where $\mathbf{a}$ is an arbitrary positional offset. We see that the offset produces a variable phase factor which correctly aligns the function without modifying the frequency amplitude, i.e., $|e^{-2\pi i\boldsymbol{\varpi}\cdot\mathbf{a}}\tilde{f}(\boldsymbol{\varpi})|=|\tilde{f}(\boldsymbol{\varpi})|$. However, when the inverse Fourier transform is applied this phase factor can have an impact on the accuracy of the discretization of f. This means that the discrepancy functions Δ_h and Δ_h^{VA} depend on the offset $\mathbf{a}$ and should therefore, in principle, be evaluated for all $\mathbf{a}\in[0,h]^3$. Even if we could do this, in practice we will not be able to control the actual offset. Therefore, we examine the definition of the inverse Fourier transform in Eq. (2) and deduce the following inequality:

$$|f(\mathbf{r})|\leq\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty}|\tilde{f}(\boldsymbol{\varpi})|d^3\boldsymbol{\varpi} \tag{20}$$

which follows from the fact that the $e^{-2\pi i\boldsymbol{\varpi}\cdot\mathbf{r}}$ term can at most return the absolute value of $\tilde{f}(\boldsymbol{\varpi})$ when all of its components are in phase. Applying this logic to the discrepancy functions

in Eqs. (17) and (18), while recalling that the last terms in these equations simply remove one of the 27 terms from the triple summations, yields the following bounding inequalities:

$$|\Delta_h(\mathbf{r})| \leq \left(\sum_{i=1}^{3}\sum_{j=1}^{3}\sum_{k=1}^{3}\int_{A_i}\int_{A_j}\int_{A_k}|\tilde{d}(\boldsymbol{\varpi})|d^3\boldsymbol{\varpi}\right) - \int_{A_2}\int_{A_2}\int_{A_2}|\tilde{d}(\boldsymbol{\varpi})|d^3\boldsymbol{\varpi},$$

$$|\Delta_h^{\mathrm{VA}}(\mathbf{r})| \leq \left(\sum_{i=1}^{3}\sum_{j=1}^{3}\sum_{k=1}^{3}\int_{A_i}\int_{A_j}\int_{A_k}|\tilde{d}(\boldsymbol{\varpi})^3\tilde{\Pi}_h(\boldsymbol{\varpi})|d^3\boldsymbol{\varpi}\right) - \int_{A_2}\int_{A_2}\int_{A_2}|\tilde{d}(\boldsymbol{\varpi})^3\tilde{\Pi}_h(\boldsymbol{\varpi})|d^3\boldsymbol{\varpi} \quad (21)$$

whose right-hand sides yield the worst case discrepancies regardless of the alignment of the discretization grid. This not only provides a more general bound on the errors associated with discretization and volume averaging but are numerically much easier to handle than Eqs. (17) and (18) due to the elimination of the oscillatory $e^{-2\pi i\boldsymbol{\varpi}\cdot\mathbf{r}}$ terms for large $\boldsymbol{\varpi}$. Typically, functions that represent physical phenomena, such as the dose distribution of a MV photon beam in a water phantom, have vanishing high frequency contributions that allow for accurate numerical computation of Eqs. (21).

Note that we can use the same approach to bound the discrepancy between the true dose distribution d and the discretization $\check{d}_h^{\mathrm{VA}}$ of the volume-averaged dose distribution function $|d-\check{d}_h^{\mathrm{VA}}|$:

$$|d(\mathbf{r})-\check{d}_h^{\mathrm{VA}}(\mathbf{r})| \leq |d(\mathbf{r})-d_h^{\mathrm{VA}}(\mathbf{r})| + |d_h^{\mathrm{VA}}(\mathbf{r})-\check{d}_h^{\mathrm{VA}}(\mathbf{r})| = |d(\mathbf{r})-d_h^{\mathrm{VA}}(\mathbf{r})| + |\Delta_h^{\mathrm{VA}}(\mathbf{r})|. \quad (22)$$

Now note that the first term on the right-hand side is the discrepancy between the true dose distribution and the volume averaged dose distribution as a function of the dimension of the voxel, h. To compute a bound on this quantity, we may substitute the difference between d and d_h^{VA}, which are given in Eqs. (15), into Eq. (20):

$$|d(\mathbf{r})-d_h^{\mathrm{VA}}(\mathbf{r})| \leq \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty}|[1-{}^3\tilde{\Pi}_h(\boldsymbol{\varpi})]\tilde{d}(\boldsymbol{\varpi})|d^3\boldsymbol{\varpi}. \quad (23)$$

We can see that the term $[1-{}^3\tilde{\Pi}_h(\boldsymbol{\varpi})]$ acts as a low pass filter on the Fourier transform of the dose distribution rather than an abrupt cutoff as seen in Eqs. (21). It is even more interesting to consider a finite value of h and ask how the two discretizations differ. Using a similar analysis as earlier, we can derive a bound on the discrepancy between the dose distribution obtained by point sampling $\check{d}_h$ and the discretization $\check{d}_h^{\mathrm{VA}}$ of the volume-averaged dose distribution function $|\check{d}_h-\check{d}_h^{\mathrm{VA}}|$:

$$|\check{d}_h(\mathbf{r})-\check{d}_h^{\mathrm{VA}}(\mathbf{r})| \leq \int_{A_2}\int_{A_2}\int_{A_2}|[1-{}^3\tilde{\Pi}_h(\boldsymbol{\varpi})]\tilde{d}(\boldsymbol{\varpi})|d^3\boldsymbol{\varpi}. \quad (24)$$

As a by-product, this leads to the intuitively obvious fact that in the limit of vanishing voxel size discrete sampling and volume averaging are equivalent. For the same reasons as given earlier, Eqs. (23) and (24) can accurately be numerically evaluated as a function of h.

B. Application to a beamlet-based dose model for IMRT FMO

FMO models for IMRT treatment planning are formulated using the assumptions of linearity (i.e., superposition and scaling) applied to many small beams or beamlets. The absorbed ionizing radiation dose received by each voxel due to all beamlets, $D(\mathbf{r})$, in such a model is then the summation of the dose distributions, $d_j(\mathbf{r})$, arising from the individual beamlets, j, weighted by beamlet fluences u_j as follows:

$$D(\mathbf{r}) = \sum_{j=1}^{N_b} d_j(\mathbf{r})u_j, \quad (25)$$

where N_b is the number of beamlets. Our goal is to estimate the impact of discretization on the cumulative dose distribution $D(\mathbf{r})$. However, it is worth investigating if the error can be deduced from the individual beamlet or if the composite is required. From our analysis of discretization errors in the previous section we can see that this question is equivalent to asking if the Fourier spectrum of $D(\mathbf{r})$ is different from the Fourier spectrum of $d(\mathbf{r})$. To answer this question we point out the linearity of the Fourier transform, i.e.:

$$\mathfrak{I}[af(\mathbf{x})+bg(\mathbf{x})] = a\tilde{f}(\boldsymbol{\varpi})+b\tilde{g}(\boldsymbol{\varpi}) \quad (26)$$

from which we can deduce that

$$\tilde{D}(\boldsymbol{\varpi}) = \sum_{j=1}^{N_b}\mu_j\tilde{d}_j(\boldsymbol{\varpi})e^{-2\pi i\boldsymbol{\varpi}\cdot\mathbf{a}_j}. \quad (27)$$

This shows that the Fourier spectrum of $D(\mathbf{r})$ is just a weighted sum of the Fourier spectra of the $d_j(\mathbf{r})$'s where a phase factor of $e^{-2\pi i\boldsymbol{\varpi}\cdot\mathbf{a}}$ has been added to account for the fact that the beamlets will have different spatial locations in the composite plan. Note, however, that the phase factor does not materially influence the error estimation from the previous section. If we now assume that changes in the different $d_j(\mathbf{r})$ due to differences in source to surface distance (SSD) and heterogeneities result in relatively small differences in a nominal dose-to-water beamlet Fourier spectrum $\tilde{d}(\boldsymbol{\varpi})$, then we may restrict our study to a single beamlet rather than the composite dose distribution. This assumption is reasonable in cases when the major interest is in dose to water-like tissues. The major gradient differences due to heterogeneities will be localized in and very near to higher and lower density regions and their interfaces to unit density regions. Due to the treatment distance, changes in SSD do not matter as much as changes in density as to first order, SSD changes only influ-

ence the central axis percent depth dose (PDD) which changes very slowly compared to the lateral beamlet profile. In fact, as we are only interested in a worst case estimate of the error we will restrict our study to the lateral profile of the dose distribution as it contains the highest possible gradients and thus the largest frequencies.

We now consider a specific functional form for our beamlet dose distributions, $d(\mathbf{r})$, which are based upon a fit to high resolution (100 μm pixel) radiochromic film measurements and have been reported elsewhere.[13,14] To compute dose as a function of Cartesian beam coordinates (with the source at the origin) where the central beam axis points along the $+z$ direction, at a radiological depth, rd, and asymmetric field size, $s_x \times s_y$, we employ a simple empirical beamlet model of the form

$$d(\mathbf{r}) = \mathrm{SPDD}(\mathrm{rd}) \times d^{\mathrm{OA}}(\mathbf{r}) \times \left(\frac{100+d}{z}\right)^2, \tag{28}$$

where we define the off-axis function

$$d^{\mathrm{OA}}(\mathbf{r}) = d_x^{\mathrm{OA}}(\mathbf{r}) \times d_y^{\mathrm{OA}}(\mathbf{r}), \tag{29}$$

where

$$d_\nu^{\mathrm{OA}}(\mathbf{r}) = \frac{1}{2}\left\{ \mathrm{erf}\left[\frac{\nu+\nu_0}{\sigma(\mathrm{rd})}\right] - \mathrm{erf}\left[\frac{\nu-\nu_0}{\sigma(\mathrm{rd})}\right]\right\} \quad \text{for } \nu = x,y, \tag{30}$$

$\nu_0 = z \times s_\nu / 100$, SPPD is a fit to the percent depth dose along the central axis with the inverse square factor, $[(100+d)/z]^2$, scaled out, and $\sigma(\mathrm{rd})$ is the standard deviation of the Gaussian in the integrand of the error function, erf, as a function of radiological depth. As the value of $\sigma(\mathrm{rd})$ increases very slowly with depth, we took it to have a worst case constant value of 2 mm corresponding to the penumbra at the depth of maximum depth, leading to a 20%–80% distance of 2.25 mm.

C. Empirical analysis of discretization error in FMO

To obtain a practical case for empirical analysis, we studied the effect of dose grid resolution on the quality of the treatment plans obtained for an IMRT head-and-neck case where the primary reason for using IMRT was to preserve salivary function while obtaining adequate target coverage to two targets with prescription doses of 73.8 and 54 Gy. In addition to salivary glands, brainstem, spinal cord, mandible, and unspecified tissue were also spared as critical structures in the treatment plans. 3D treatment planning data were exported via DICOM from a commercial patient imaging and anatomy segmentation system (VoxelQ, Philips Medical Systems) into the University of Florida Optimized Radiation Therapy (UFORT) treatment planning system where a previously published FMO model was used to solve for globally optimal fluences for the same case using 1, 2, 3, 4, 5, and 6 mm isotropic point dose-grid spacings.[13] Each solution was then evaluated using both the modeling resolution (2, 3, 4, 5, and 6 mm) and the 1 mm dose grid.

Given that the dose model employed is analytic, no fluence map filtering was applied (though a systematic fluence map intensity rounding at 5% resolution is performed and the model is tuned to account for this), and the optimization algorithm produces globally optimal solutions to a numerically defined precision the only significant errors that can be observed in this comparative analysis are discretization errors. These errors directly influence the beamlet weights chosen by the optimization due to aliasing, i.e., the model does not have adequate information to form an optimal solution. The aliasing errors in the modeled fluence are directly proportional to the dose errors observed as dose is a linear function of fluence [see Eq. (25)]. Unless the dose is reviewed at higher resolution this aliasing effect is hidden from the reviewer.

Dose-volume histograms (DVHs) were estimated by simply binning point doses computed on the isotropic dose grids specified in each model into 1 Gy dose bins. The use of 1 Gy dose bins was considered adequate to represent any clinical significant features in the computed dose distributions. Based on a comparison of the apparent and actual dose distributions obtained using these resolutions, a multiresolution model was also developed and solved where isotropic 2, 4, and 6 mm point dose grids were employed in the targets, critical structures, and unspecified tissue, respectively. This multiresolution model was then compared to the 2 mm model. Finally, a two-dimensional (2D) histogram was generated of the norm of the cumulative dose distribution gradient, binned in 1 (Gy/mm) increments, and the dose discrepancy between the 4 and 1 mm treatment plan dose distributions.

D. Comparison of theory and empirical results

1. Theoretical errors due to point sampling

Worst case error estimates were evaluated for 1–6 mm point dose-grid spacings using the expression for $|\Delta_h(\mathbf{r})|$ in Eqs. (21), as the UFORT system employs analytic point sampling in its dose calculation algorithm. The Fourier transform of the lateral profile, $d_\nu^{\mathrm{OA}}(\mathbf{r})$:

$$\tilde{d}_\nu^{\mathrm{OA}}(\varpi) = 2\nu_0 e^{-\pi^2\sigma^2\varpi^2}\mathrm{sinc}(2\varpi\nu_0) \tag{31}$$

was found using the derivative theorem of Fourier transforms (see, e.g., p. 124 in Ref. 11) as presented in the Appendix and yields the following discrepancy bound along the x axis (and by symmetry the y axis)

$$|\Delta_h(\mathbf{r})| \leq 2\int_{+1/2h}^{+\infty} |2x_0 e^{-\pi^2\sigma^2\kappa^2}\mathrm{sinc}(2\kappa x_0)| d\kappa. \tag{32}$$

The right-hand size of Eq. (31) was evaluated for all six dose-grid spacings using a commercial numerical analysis software package (Mathematica 5, Wolfram Research Inc., Champaign, IL).

2. Theoretical errors due to volume averaging

Combining Eqs. (21), (23), and (31), and substituting them into Eq. (22) we obtain an expression for the error caused solely by volume averaging

$$|d(\mathbf{r}) - \check{d}_h^{\mathrm{VA}}(\mathbf{r})| \leq \int_{-\infty}^{+\infty} |2x_0 e^{-\pi^2\sigma^2\kappa^2} \mathrm{sinc}(2\kappa x_0) \times [1 - \mathrm{sinc}(h\kappa)]| d\kappa + 2\int_{+1/2h}^{+\infty} |2x_0 e^{-\pi^2\sigma^2\kappa^2} \mathrm{sinc}(2\kappa x_0) \times \mathrm{sinc}(h\kappa)| d\kappa. \quad (33)$$

Because this expression predicts a different error bound than the bound on point sampling, it is interesting when attempting to compare simplistic analytic dose models against more elegant stochastic dose models as has recently been performed by Jeraj *et al.*[15] The results of this analysis allowed us to evaluate when such comparisons can be influenced by the voxel size. Worst case error estimates were evaluated for 1–6 mm isotropic voxel dose grids be evaluating the right-hand side of Eq. (32) using the commercial numerical analysis software.

III. RESULTS

A. Empirical analysis of discretization error in FMO

The treatment plans obtained using 2, 3, 4, 5, and 6 mm isotropic dose grids for the FMO were reviewed using both the FMO grid and a 1 mm dose grid. Only the 1 and 2 mm plans were in agreement for all dose-volume criteria to better than 2% for all targets and structures. Empirically, this convergence of results suggests that a 2 mm isotropic dose grid is sufficient to provide accurate dose information. The most clinically significant differences were found for the high-dose target coverage when the grid spacing was 3 mm or larger. In these cases, targets always *appeared* to be well covered as seen in Fig. 1(a) where high-dose target DVHs were computed using the reported resolution, but lost both coverage and homogeneity *in reality* when the DVHs were recomputed using a 1 or 2 mm grid, as seen in Fig. 1(b). High-dose target coverage dropped from the planned and apparent 95% coverage to an actual 90%, 86.5%, 79%, and 67% coverage for 3, 4, 5, and 6 mm grids, respectively, while the target coverage was unchanged for the 2 mm grid. The target dose inhomogeneity increased steadily from +11% and −3% to +20% and −22% as the resolution spanned 3–6 mm. The low dose target also suffered coverage and homogeneity losses, though approximately as half as much as the high-dose target. In general, critical structures suffered from similar errors as targets but upon treatment plan review these errors were found to be less clinically significant. Therefore, good clinical sparing was maintained by the model having grids of 4 mm or less. Similar to the case for targets, critical structures always *appeared* to have better plans, as seen in Fig. 2(a) where DVHs were computed for the spinal cord using the reported resolution, but lost sparing *in reality* when the DVHs were recomputed using a 1 or 2 mm grid. However, the resolution required to preserve sparing was half that required by targets. The sparing of the unspecified tissue required even less resolution than other structures, with adequate sparing obtained with a grid spacing as large as 6 mm.

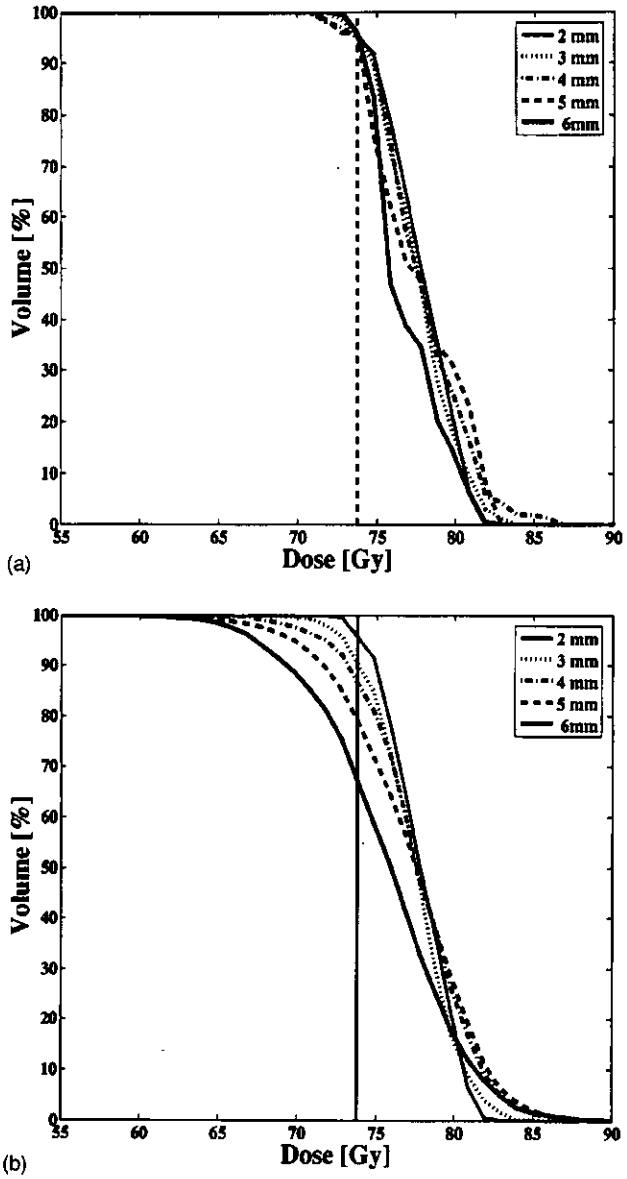
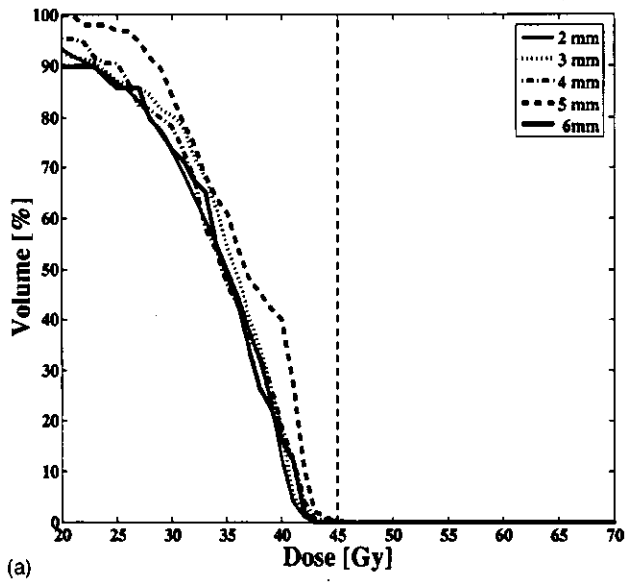


FIG. 1. An overlay of high-dose target DVHs computed using the (a) the native FMO model resolutions of 2, 3, 4, 5, and 6 mm to demonstrate the assumed plan quality and (b) the effective Nyquist–Shannon limited resolution of 1 mm to demonstrate the actual plan quality.

Use of the 2, 4, and 6 mm multiresolution model, where grid spacings of 2, 4, and 6 mm were applied to targets, critical structures, and unspecified tissues, respectively, produced excellent results as seen in Fig. 3. The 2D histogram of cumulative dose gradient versus the dose error due to point sampling discretization produced a 2D distribution consistent with a normal distribution having a most probable value of zero. This result indicates that there were no correlations between the gradient of the cumulative dose distribution and the observed discretization errors.

B. Comparison of theory and empirical results

Evaluation of the right-hand side of Eq. (32) provided the worst case error predictions for isotropic point sampling



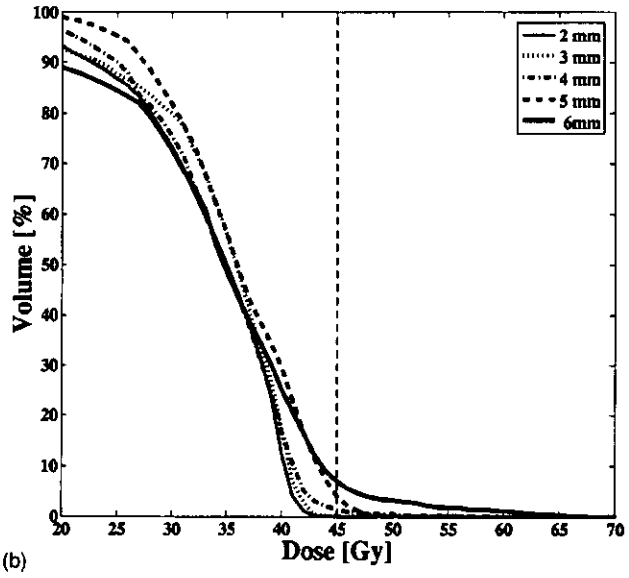


FIG. 2. An overlay of spinal cord DVHs computed using the (a) the native FMO model resolutions of 2, 3, 4, 5, and 6 mm to demonstrate the assumed plan quality and (b) the effective Nyquist–Shannon limited resolution of 1 mm to demonstrate the actual plan quality.

grids of 1.0, 1.5, 2.0, 2.5, 3.0, 3.5, 4.0, 5.0, and 6.0 mm. This result, seen in the second column of Table I, correlated well with the increases in the high-dose target dose heterogeneity. The theory predicted that minimizing the worst case errors at the maximum dose value to less than 2% would require an isotropic point grid spacing of between 2.5 and 3 mm or smaller. The increase in error with increasing grid spacing can also be observed. These data were consistent with the empirical results that a 2 mm grid provided better than 2% agreement, while a 3 mm grid could produce errors more than twice as large. Evaluation of the right-hand side of Eq. (33) provided the worst case bound on the difference between the true dose distribution and the discretized volume-averaged distribution. This result, seen in the third column of Table I, demonstrated that error due to volume averaging and

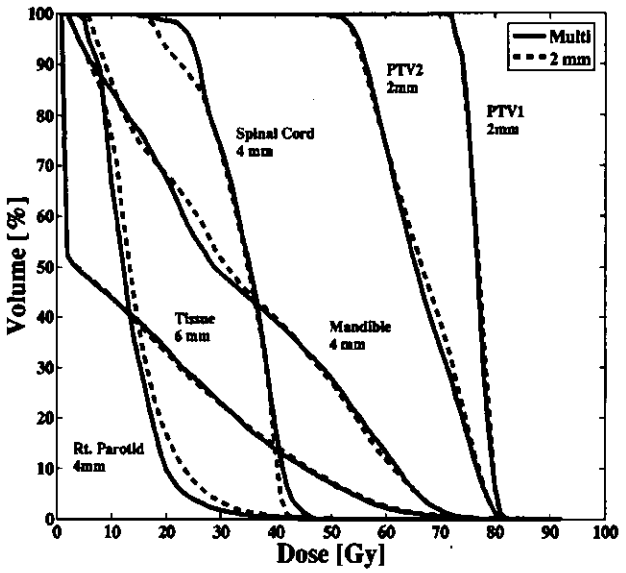


FIG. 3. A comparison of the DVHs from the 2 mm FMO treatment plan (dashed lines) and the 2, 4, and 6 mm multiresolution model FMO treatment plan (solid lines).

sampling can be larger than the error due to point sampling alone and increases in a similar fashion. Here, the theory predicted that worst case volume averaging errors at the maximum dose value of less than 2% would require voxel dimension of between 1.5 and 2.0 mm.

IV. DISCUSSION AND CONCLUSIONS

In this article, we have presented a theoretical and empirical analysis of the errors associated with approximating IMRT dose distributions by point sampling and discretized volume averaging discretizations. The information theory based Fourier analysis of this problem allowed us to determine that the general IMRT problem can be effectively studied by examining the resolution requirements of a single beamlet. The theory produced analytic integrals that bounded worst case aliasing errors that can occur regardless of the location and orientation of the dose grid with respect to the dose distribution. The predictions of this theory were found to be in reasonable agreement with empirical results obtained

TABLE I. Discretization errors at the maximum 6 MV beamlet dose value as a function of dose grid spacing.

h (mm)	Max point sampling error at max value (%)	Max discretized volume averaging error at max value (%)
1.0	0.000 074 4	0.847
1.5	0.0616	1.93
2.0	0.575	3.59
2.5	1.66	6.04
3.0	3.30	8.46
3.5	7.01	13.2
4.0	10.6	16.1
5.0	13.6	20.9
6.0	15.8	25.4

by solving a linear-programming based IMRT FMO model to global optimality. The attainment of worst case estimates was attributed to the fact that the optimization takes advantage of aliasing to produce an optimal plan. The empirical observation that no correlation existed between the cumulative dose distribution gradient and the observed discretization errors supported the fact that it is the information contained in the beamlets of the IMRT dose model that control the error in the cumulative plan. It was predicted that an isotropic dose grid with 2.5 mm spacing is sufficient to prevent dose errors larger than two percent. While not supported by the theory, it is noted that in practice this resolution was mostly needed in high-dose target regions. In light of the empirical results, we have suggested a multiresolution 2, 4, and 6 mm spacing model where these spacings are applied to targets, structures, and tissue, respectively. This model allowed a significant reduction of the number of variables and constraints required to solve an IMRT FMO problem while providing accurate dose and DVH information. Treatment plans should still be reviewed with higher resolution in all structures.

The general result of this study, that IMRT FMO places a higher requirement on dose grid resolution than conventional radiation therapy, is not surprising given the reliance of IMRT on the superposition of the penumbra of many apertures to produce a desired dose distribution. While 3–4 mm grid resolution has been previously deemed adequate for conformal therapy treatment planning, it would appear that a 2.5 mm dose grid is required at least in regions of high dose for IMRT FMO for beams having a similar penumbra. In general, a 2.5 mm grid provides an effectively Nyquist–Shannon-limited sampling of the 6 MV dose distributions with a 20%–80% penumbra distance of 2.25 mm. More accurate modeling of heterogeneities might introduce higher gradients in the dose distribution that could require higher resolution in some cases. Further study into discretization errors in the presence of tissue or phantom heterogeneities should be performed numerically in 3D using high-resolution dose grids and more accurate dose models. The potential for errors between point-sampled and volume-averaged dose distributions should be considered when comparing different approaches to accounting for heterogeneities in IMRT optimization. Finally, an investigation into the more conformal modalities of proton and heavy ion therapy should be pursued.

ACKNOWLEDGMENT

This research was supported in part by NIH Grant No. R01 CA100636 awarded by the National Cancer Institute.

APPENDIX: DERIVATION OF THE FOURIER TRANSFORM OF THE DIFFERENCE OF GENERALIZED erf FUNCTIONS

We now derive the Fourier transform of the difference of two offset and scaled erf functions. We consider the general functional form

$$p(\nu)=\frac{1}{2}\left[\operatorname{erf}\left(\frac{\nu+\nu_0}{\sigma}\right)-\operatorname{erf}\left(\frac{\nu-\nu_0}{\sigma}\right)\right]. \tag{A1}$$

This functional form was found to fit the measured lateral profile of small beams measured using high spatial resolution radiochromic film. We first seek the Fourier transform of the erf function

$$\operatorname{erf}(x)=\frac{2}{\sqrt{\pi}}\int_0^x e^{-t^2}dt. \tag{A2}$$

To this end, first note that the Fourier transform of the derivative of a function is related to the Fourier transform of the original function by

$$\mathfrak{I}[f'(x)]=2\pi i\varpi\tilde{f}(\varpi) \tag{A3}$$

which implies that

$$\widetilde{\operatorname{erf}}(\varpi)=\frac{1}{2\pi i\varpi}\mathfrak{I}[\operatorname{erf}'(x)]. \tag{A4}$$

Then, note that the derivative of the erf function is a Gaussian function

$$\operatorname{erf}'(x)=\frac{2}{\sqrt{\pi}}e^{-x^2} \tag{A5}$$

and that a Gaussian is its own Fourier transform,

$$\mathfrak{I}[e^{-\pi x^2}]=e^{-\pi\varpi^2}. \tag{A6}$$

Now we point out that the Fourier transform has the following scaling property:

$$\mathfrak{I}[f(ax)]=\frac{1}{|a|}\tilde{f}\left(\frac{\varpi}{a}\right) \tag{A7}$$

which yields that

$$\begin{aligned}\mathfrak{I}[\operatorname{erf}'(x)]&=\mathfrak{I}\left[\frac{2}{\sqrt{\pi}}e^{-x^2}\right]\\&=\frac{2}{\sqrt{\pi}}\mathfrak{I}[e^{-x^2}]\\&=\frac{2}{\sqrt{\pi}}\mathfrak{I}[e^{-\pi(x/\sqrt{\pi})^2}]=2e^{-\pi(\varpi\sqrt{\pi})^2}=2e^{-\pi^2\varpi^2}.\end{aligned} \tag{A8}$$

Combining this result with Eq. (A4), we then obtain that

$$\widetilde{\operatorname{erf}}(\varpi)=\frac{1}{\pi i\varpi}e^{-\pi^2\varpi^2}. \tag{A9}$$

Next, we will determine the Fourier transforms of

$$\operatorname{erf}\left(\frac{\nu+\nu_0}{\sigma}\right) \quad \text{and} \quad \operatorname{erf}\left(\frac{\nu-\nu_0}{\sigma}\right) \tag{A10}$$

in two steps. First, apply again the scaling property from Eq. (A7) to obtain that

$$\Im\left[\operatorname{erf}\left(\frac{\nu}{\sigma}\right)\right] = \sigma \times \widetilde{\operatorname{erf}}(\varpi\sigma) = \sigma\frac{1}{\pi i\varpi\sigma}e^{-\pi^2(\varpi\sigma)^2}$$

$$= \frac{1}{\pi i\varpi}e^{-\pi^2\varpi^2\sigma^2}. \quad \text{(A11)}$$

Next, recall that the impact of applying a shift or offset to a function on its Fourier transform was described in Eq. (19) earlier. Applying this offset property yields

$$\Im\left[\operatorname{erf}\left(\frac{\nu+\nu_0}{\sigma}\right)\right] = e^{-2\pi i\varpi\nu_0} \times \frac{1}{\pi i\varpi}e^{-\pi^2\varpi^2\sigma^2},$$

$$\Im\left[\operatorname{erf}\left(\frac{\nu-\nu_0}{\sigma}\right)\right] = e^{2\pi i\varpi\nu_0} \times \frac{1}{\pi i\varpi}e^{-\pi^2\varpi^2\sigma^2}. \quad \text{(A12)}$$

Finally, substituting Eqs. (A12) into Eq. (A1) we obtain

$$\Im[p(\nu)] = \Im\left\{\frac{1}{2}\left[\operatorname{erf}\left(\frac{\nu+\nu_0}{\sigma}\right) - \operatorname{erf}\left(\frac{\nu-\nu_0}{\sigma}\right)\right]\right\}$$

$$= \frac{1}{2}\Im\left[\operatorname{erf}\left(\frac{\nu+\nu_0}{\sigma}\right)\right] - \frac{1}{2}\Im\left[\operatorname{erf}\left(\frac{\nu-\nu_0}{\sigma}\right)\right]$$

$$= \frac{1}{2\pi i\varpi}e^{-\pi^2\varpi^2\sigma^2} \times (e^{-2\pi i\varpi\nu_0} - e^{2\pi i\varpi\nu_0})$$

$$= \frac{1}{2\pi i\varpi}e^{-\pi^2\varpi^2\sigma^2} \times [2i\sin(2\pi\varpi\nu_0)]$$

$$= e^{-\pi^2\varpi^2\sigma^2} \times \frac{\sin(2\pi\varpi\nu_0)}{\pi\varpi}$$

$$= 2\nu_0 e^{-\pi^2\sigma^2\varpi^2}\operatorname{sinc}(2\varpi\nu_0) = \tilde{p}(\varpi). \quad \text{(A13)}$$

[a]Author to whom correspondence should be addressed. Electronic mail: dempsey@ufl.edu

[1]A. Niemierko and M. Goitein, "The influence of the size of the grid used for dose calculation on the accuracy of dose estimation," Med. Phys. **16**, 239–247 (1989).
[2]C. W. Smith, D. Morrey, and K. Gray, "The influence of grid size on accuracy in radiotherapy dose plotting," Med. Phys. **17**, 135–136 (1990).
[3]X.-Q. Lu and L. M. Chin, "Sampling techniques for the evaluation of treatment plans," Med. Phys. **20**, 151–161 (1993).
[4]A. Jackson, R. Mohan, and B. Baldwin, "Comments on 'Sampling techniques for the evaluation of treatment plans,'" Med. Phys. **20**, 1375–1376 (1993).
[5]A. Niemierko and M. Goitein, "Comments on 'Sampling techniques for the evaluation of treatment plans,'" Med. Phys. **20**, 1377–1380 (1993).
[6]A. Niemierko and M. Goitein, "The use of variable grid spacing to accelerate dose calculations," Med. Phys. **16**, 357–366 (1989).
[7]A. Niemierko and M. Goitein, "Random sampling for evaluating treatment plans," Med. Phys. **17**, 753–762 (1990).
[8]T. Bortfeld, U. Oelfke, and S. Nil, "What is the optimum leaf width of a multileaf collimator?," Med. Phys. **27**, 2494–2502 (2000).
[9]H. Nyquist, "Certain topics in telegraph transmission theory," Trans. Am. Inst. Electr. Eng. **47**, 617–644 (1928).
[10]C. E. Shannon and W. Weaver, *The Mathematical Theory of Communication* (University of Illinois Press, 1998), 144 pp.
[11]R. Bracewell, *The Fourier Transform and Its Applications*, 3rd ed. (McGraw-Hill, New York, 1999).
[12]K. R. Kase and W. R. Nelson, "Concepts of Radiation Dosimetry," Stanford Linear Accelerator Center (SLAC) Report, SLAC-R-0153, 1972, pp. 4–6 (http://www.slac.stanford.edu/pubs/).
[13]J. W. Sohn, J. F. Dempsey, T. S. Suh, and D. A. Low, "Analysis of various beamlet sizes for IMRT with 6 MV photons," Med. Phys. **30**, 2432–2439 (2003).
[14]H. E. Romeijn, R. K. Ahuja, J. F. Dempsey, A. Kumar, and J. G. Li, "A novel linear programming approach to fluence map optimization for intensity modulated radiation therapy treatment planning," Phys. Med. Biol. **48**, 3521–3542 (2003).
[15]R. Jeraj, P. J. Keall, and J. V. Siebers, "The effect of dose calculation accuracy on inverse treatment planning," Phys. Med. Biol. **47**, 391–407 (2002).

INSTITUTE OF PHYSICS PUBLISHING PHYSICS IN MEDICINE AND BIOLOGY

Phys. Med. Biol. 48 (2003) 3521–3542 PII: S0031-9155(03)62422-X

A novel linear programming approach to fluence map optimization for intensity modulated radiation therapy treatment planning

H Edwin Romeijn[1], Ravindra K Ahuja[1], James F Dempsey[2], Arvind Kumar[1] and Jonathan G Li[2]

[1] Department of Industrial and Systems Engineering, University of Florida, Gainesville, FL 32611-6595, USA
[2] Department of Radiation Oncology, University of Florida, Gainesville, FL 32610-0385, USA

E-mail: romeijn@ise.ufl.edu, ahuja@ufl.edu, dempsey@ufl.edu, arvind22@ufl.edu and lijg@ufl.edu

Received 18 April 2003, in final form 1 September 2003
Published 10 October 2003
Online at stacks.iop.org/PMB/48/3521

Abstract

We present a novel linear programming (LP) based approach for efficiently solving the intensity modulated radiation therapy (IMRT) fluence-map optimization (FMO) problem to global optimality. Our model overcomes the apparent limitations of a linear-programming approach by approximating any convex objective function by a piecewise linear convex function. This approach allows us to retain the flexibility offered by general convex objective functions, while allowing us to formulate the FMO problem as a LP problem. In addition, a novel type of partial-volume constraint that bounds the tail averages of the differential dose–volume histograms of structures is imposed while retaining linearity as an alternative approach to improve dose homogeneity in the target volumes, and to attempt to spare as many critical structures as possible. The goal of this work is to develop a very rapid global optimization approach that finds high quality dose distributions. Implementation of this model has demonstrated excellent results. We found globally optimal solutions for eight 7-beam head-and-neck cases in less than 3 min of computational time on a single processor personal computer without the use of partial-volume constraints. Adding such constraints increased the running times by a factor of 2–3, but improved the sparing of critical structures. All cases demonstrated excellent target coverage (>95%), target homogeneity (<10% overdosing and <7% underdosing) and organ sparing using at least one of the two models.

(Some figures in this article are in colour only in the electronic version)

0031-9155/03/213521+22$30.00

1. Introduction

Many methods of optimization have been proposed or used in attempts to solve the beamlet-intensity or fluence-map optimization (FMO) problem for mega-voltage photon-beam intensity modulated radiation therapy (IMRT) (Webb 2001, Shepard *et al* 1999). We conceptually break the task of solving an optimization problem into two components: (i) model formulation and (ii) algorithm selection and implementation. A good deal of discussion regarding model formulation has appeared in the medical physics literature. Most of this work is focused on the topologic properties of FMO models, such as the unimodality or existence of non-global local minima (Alber *et al* 2002, Deasy 1997, Llacer *et al* 2003, Wu and Mohan 2002). There are many competing models based on both 'physical' (Bortfeld 1999) and 'biological' (Wu *et al* 2002, 2003, Niemierko *et al* 1992, Niemierko 1997, Mavroidis *et al* 2001, Kallman *et al* 1992, Jones and Hoban 2000, Alber and Nusslin 1999) objective functions and constraints, possibly with many more under development. While it has been pointed out that several widely employed model formulations for FMO are in fact unimodal (Deasy 1997), other FMO formulations, in particular those that employ dose–volume constraints, have been shown to be nonconvex, which may lead to the existence of multiple local minima. Due to the combinatorially large number of trade-offs that can be made between covering radiotherapy targets and sparing critical tissues, it is difficult to establish an *a priori* or fundamental basis for model preference. It would appear that many presented models can produce high quality solutions. A concern exists when comparing these models, as the results can depend on the algorithms used in solving them. Little attention has been paid to algorithm selection and implementation for FMO in medical physics. This second step is quite significant, and it is in fact a major focus of the optimization theory literature. It is well known, for example, that not all local-search algorithms are guaranteed to solve even a unimodal model (a model with a single local minimum) to optimality without needing essentially unlimited computational precision and time. Thus, the mere fact that several widely employed model formulations for FMO are unimodal does not guarantee that solutions to these formulations found by a given implementation of an algorithm will be optimal. If a globally optimal solution cannot be guaranteed, it is impossible to decide whether the model formulation used is unsuitable, or that the selected algorithm simply cannot find the best solution for that formulation. Even when optimal solutions can be guaranteed, the chosen algorithm may be highly inefficient. For example, when designed and implemented correctly, the simulated annealing algorithm (Webb 1994, 1995) is guaranteed to find a globally optimal solution given unlimited time (Hajek 1988, Romeijn and Smith 1994), but the convergence to the globally optimal solution often tends to be far too slow for practical purposes. Another confounding factor is that many algorithms have no means to recognize the fact that a globally optimal solution has been found, should it arrive there. In developing a strategy to produce a robust method for solving the FMO problem to optimality, we believe that the best approaches will need to combine a unimodal model formulation with an efficient algorithm that can solve the problem to global optimality and recognize that global optimality has been achieved. Given that the definition of global optimality is model dependent, it is clear that applying such an approach to different FMO model formulations will allow for a true evaluation of the efficacy of these models.

In a comprehensive review of the radiation therapy optimization literature, Shepard *et al* (1999) surveyed many techniques previously used. This review identified three classes of mathematical programming models for the FMO problem: linear programming (LP), mixed-integer LP (MILP) and nonlinear programming (NLP). Even before the advent of IMRT model development, LP models that imposed simple upper and lower dose constraints with (Bahr *et al* 1968, Langer and Leong 1987, Morrill *et al* 1991b) or without (Censor *et al* 1988) linear

objective functions were proposed in the literature for conformal treatment plan optimization. However, LP models were deemed inappropriate (Shepard *et al* 1999) to adequately solve the FMO problem due to the apparent limitations in formulating suitable linear objective functions and constraints that would provide high quality treatment plans. Therefore, despite the availability of very efficient algorithms that can find and recognize globally optimal solutions to very large-scale LP problems, LP has found little use in radiation therapy to date. While it was pointed out that MILP can be used to explicitly impose dose–volume constraints, it was deemed to suffer from the same limitations as LP. On the other hand, NLP, and in particular convex programming, was considered a promising tool since it provides sufficient flexibility to allow the formulation of suitable models for FMO.

The use of quadratic penalty functions and dose–volume constraints in IMRT treatment planning leads to a NLP formulation of the FMO problem. Most algorithms proposed for these formulations are heuristic in nature. Common approaches are local search methods such as the conjugate gradient method (Shepard *et al* 2000, Xing *et al* 1998) and metaheuristic techniques such as simulated annealing (Webb 1994, 1995, Morrill *et al* 1991a) and simulated dynamics (Hou and Wang 2001, Hou *et al* 2003). Many commercially available IMRT treatment planning systems employ dose–volume constraints and use such approaches (Carol *et al* 1997, Fogliata *et al* 2003, Lof 2000). All such approaches suffer from the drawback that, even when a suitable mathematical problem formulation can be found, there is no guarantee that the best solution with respect to this formulation will be found in a reasonable amount of computational time. In fact, our clinical experience shows that such algorithms occasionally get trapped in local minima in the search space, corresponding to inferior treatment plans. This situation frustrates attempts by the treatment planner to achieve a good treatment plan by logically adjusting the model input parameters. Further, these heuristics often have no means to determine within a reasonable amount of computation time whether a given solution is indeed optimal.

Recently, LP-based multi-criteria optimization (Hamacher and Küfer 2002) and MILP (Shepard *et al* 1999, Bednarz *et al* 2002, Langer *et al* 1990, 1996, Lee *et al* 2003) models have been proposed for FMO. These methods avoid the use of heuristics and have the means to determine if a given solution is indeed optimal. The multi-criteria optimization approach deals with the trade-offs between several conflicting goals that inevitably need to be made when designing a radiation therapy treatment plan, by providing a collection of solutions from which the clinician must then make a choice (note that the decision of what is optimal is left to the human user, yet once chosen that solution is declared optimal). This approach thus attempts to overcome the limitations of LP by solving many simple problems to provide an array of candidate solutions to the user. The candidate solutions are chosen to be Pareto-efficient solutions, having the property that no evaluation criterion can be improved without deteriorating any of the conflicting criteria. While this method is unimodal and can evaluate global optimality, populating the Pareto-efficient frontier with thousands of solutions makes it computationally inefficient. As mentioned above, the MILP models have gained attention due to their ability to add flexibility by incorporating dose–volume constraints. The major disadvantage of multi-criteria optimization and MILP, however, is that the typical problem size of clinically relevant problems prohibits finding an optimal solution to such models in an acceptable amount of computational time.

In this paper, we present a novel LP model for solving the FMO problem in IMRT treatment planning. Our approach to the FMO problem overcomes the apparent limitations of LP by using linearizations of nonlinear convex penalty functions. In addition, we employ a new type of dose–volume constraint that bounds the mean value of the tail of the differential dose–volume histogram of a structure. This model satisfies our criteria for a robust FMO

model by retaining linearity, and thereby unimodality and efficient solvability of the problem. In addition, solution procedures for LP are able to recognize that an obtained solution is indeed optimal.

2. Methods

2.1. Models

To formulate our FMO model, we assume that the targets and critical structures (together simply referred to as structures) are irradiated using a predetermined set of beams, each corresponding to a particular beam angle. For each beam angle, we decompose or discretize the aperture of this beam into small beamlets (typically of size 1×1 cm^2). The value associated with each beamlet, sometimes called a bixel, represents the intensity (or, more correctly, fluence (Webb and Lomax 2001)) of the corresponding beamlet.

We denote the set of all beamlets by N. The set of all structures is denoted by S, with targets corresponding to S^1 and critical structures to S^2 (clearly, $S = S^1 \cup S^2$ and $S^1 \cap S^2 = \emptyset$). In practice, each of the structures $s \in S$ is discretized into a finite set V_s of voxels. The set of all voxels will be denoted by $V = \cup_{s \in S} V_s$. Note that the sets V_s are not necessarily disjoint. For instance, if a target has invaded a critical structure, there will be an overlap between these structures. For each voxel $j \in V$, we also define a dominant structure $s_j \in S$, based on a priority list of all structures. Finally, let D_{ij} denote the dose received by voxel j from beamlet i at unit intensity.

The core task in FMO is to find the optimal values of the beamlet intensities (bixel values) for all beamlets. We therefore define decision variables representing the intensity of beamlet i by u_i ($i \in N$), and the decision vector of all beamlet intensities by $\vec{u} \in \mathbb{R}^{|N|}$.

2.1.1. A linear programming formulation without dose–volume constraints. Constraints are usually applied to the absorbed ionizing radiation dose received by each voxel. We therefore express the dose using a beamlet model where dose is a linear function of the beamlet intensities $\vec{u}$ as follows:

$$D_j(\vec{u}) = \sum_{i \in N} D_{ij} u_i \qquad j \in V. \tag{1}$$

Beamlet dose models are inherently linear and are widely used to solve the FMO problem. This provides further motivation for formulating the FMO problem as a LP problem. Clearly, within the LP framework, it is straightforward to incorporate upper and lower bounds on the dose received by each voxel while retaining linearity of the model (Shepard *et al* 1999, Langer *et al* 1990). Typically, the same upper and lower bounds will apply to all voxels that lie within a given structure:

$$L_{s_j} \leqslant D_j(\vec{u}) \leqslant U_{s_j} \qquad j \in V \tag{2}$$

where L_{s_j} and U_{s_j} denote the lower and upper bounds on the dose to the structure that is dominant for voxel j, respectively. Note that, for each voxel, we only impose the lower and upper bounds corresponding to its dominant structure. This avoids conflicting bounds that would result from incorporating bound constraints for all structures containing a particular voxel. For example, for voxels belonging to a critical structure as well as a target, it is often impossible to simultaneously satisfy bounds representing critical structure sparing as well as bounds representing target coverage. Clearly, the lower bound corresponding to critical structures is equal to 0, i.e. $L_s = 0$ for all $s \in S^2$. In the optimization model to be

formulated, we will distinguish between the dose received by voxel j as a function of the beamlet intensities, $D_j(\vec{u})$, and the corresponding decision variable Δ_j.

In practice, clearly not all plans satisfying these bound constraints are equally desirable. To distinguish between feasible plans, we add a suitable objective function to the formulation. In particular, for each structure we will specify a penalty function of the dose received by the voxels in that structure. Denoting the penalty function for structure s by F_s, we obtain the following objective function to be minimized:

$$\sum_{j \in V} F_{s_j}(\Delta_j). \tag{3}$$

Typical linear penalty functions that have been proposed in the literature have a value proportional to the dose received by all voxels, or only by voxels in critical structures (Rosen *et al* 1991). A more sophisticated objective function that can be formulated using LP is obtained when the penalty is equal to the absolute deviation of the dose received from a prescription or tolerance dose (Shepard *et al* 1999). Clearly, lower doses in critical structures are always preferred to higher doses, so for voxels in these structures the penalty function is chosen to be one sided, i.e. only surpluses over a tolerance dose are penalized in the objective function. Since large deviations from the prescription or tolerance dose are considered to be much more important than small deviations, a more frequently used alternative is to use a weighted least squares objective function. Figure 1 illustrates smooth and piecewise linear (PWL) convex dose penalty functions for a target (with prescription dose 70 Gy) and a critical structure (with tolerance dose 30 Gy). To retain linearity of the optimization problem, in this paper we propose to consider PWL approximations to convex penalty functions. Note that we have assumed that the penalty functions are the same for all voxels in a particular structure. However, in general we may relax this assumption by allowing for an individual penalty function for each voxel without increasing the size of the problem. Note that this may change the structure of the model, so the effect on the efficiency would have to be determined by further study.

When all penalty functions are PWL and convex, our problem can be formulated as a pure LP problem, which we will refer to as the LP problem with PWL objective function: $\mathrm{LP}_{\mathrm{PWL}}$. To this end, note that we can write the PWL penalty functions (assuming they have finitely many segments) as

$$F_{s_j}(\Delta_j) = \max_{k=1,\ldots,m_{s_j}} \left\{ a_k^{s_j} \Delta_j + b_k^{s_j} \right\} \tag{4}$$

for suitably chosen slopes a_k^s and intercepts b_k^s ($s \in S$), and where k denotes the number of linear segments defining the PWL function (see figure 1). We then introduce artificial variables t_j, which represent the penalty for the dose received by voxel j, for all $j \in V$. It can then be shown that the optimization problem of minimizing $\sum_{j \in V} F_{s_j}(\Delta_j)$ subject to our constraints is equivalent to the LP problem

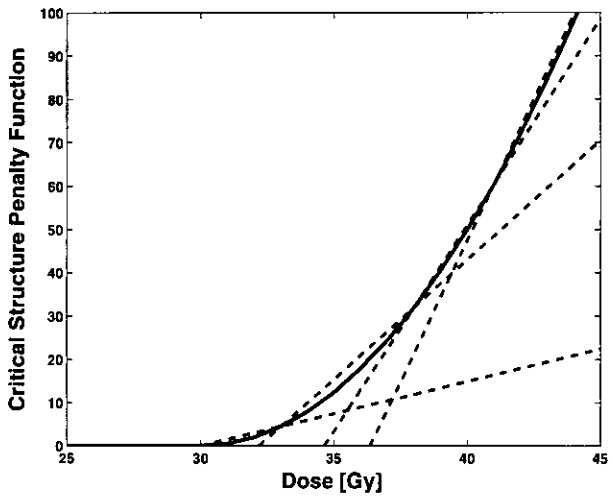
$$\text{minimize} \sum_{j \in V} t_j \tag{5}$$

subject to

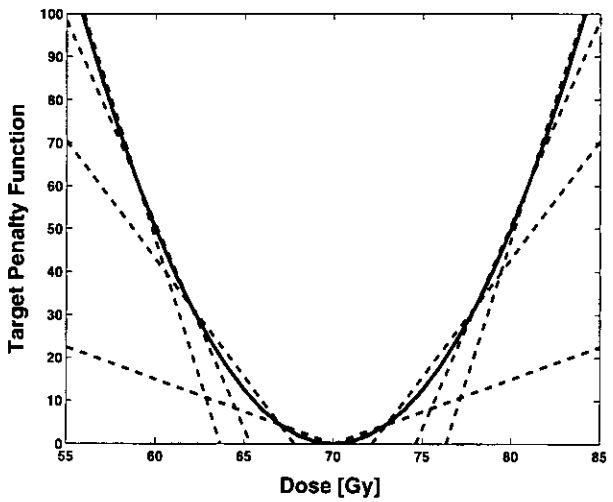
$$\Delta_j = \sum_{i \in N} D_{ij} u_i \qquad j \in V \tag{6}$$

$$a_k^{s_j} \Delta_j + b_k^{s_j} \leqslant t_j \qquad k = 1, \ldots, m_{s_j} \quad j \in V \tag{7}$$

$$L_{s_j} \leqslant \Delta_j \leqslant U_{s_j} \qquad j \in V \tag{8}$$



(a)



(b)

Figure 1. Convex voxel-based penalty functions and PWL approximations thereof for (a) a critical structure and (b) a target.

$$u_i \geqslant 0 \qquad i \in N \tag{9}$$

$$t_j, \Delta_j \geqslant 0 \qquad j \in V. \tag{10}$$

We note that there is an alternative method for modelling a PWL convex objective function in a LP model. In this method, the points of intersection of the approximating lines and the nonlinear objective function demark ranges where artificial variables are defined. The dose received by a voxel is then the sum of such artificial variables (see, e.g., Murty (1976)). Mathematically, the two methods are equivalent, but the resulting formulations may exhibit different computational efficiencies.

Unfortunately, the lower and upper bound constraints are, in practice, often conflicting. To avoid infeasibility of the optimization problem, we may allow some or all of these constraints to be violated. We can easily incorporate a penalization of violations of the lower and upper bounds on the dose received by each voxel in the penalty functions F_s, by adding very steep segments corresponding to shortfalls of a lower bound and surpluses above a upper bound, respectively.

The main difference between our model and LP formulations of the FMO problem that have previously been proposed is our choice of objective function. As Shepard *et al* (1999) noted, convex objective functions provide a high degree of modelling flexibility. Our approach is then to approximate any convex objective function by a PWL convex approximation. Such approximations retain the flexibility offered by general convex objective functions, while allowing us to formulate the problem as a LP problem.

Our $\mathrm{LP}_{\mathrm{PWL}}$ model, however, has the apparent drawback that it does not capture dose–volume constraints. As mentioned above, traditional dose–volume constraints are inherently nonlinear by nature and their exact handling causes nonconvexity in the problem formulation, for example, by using integer decision variables. In the next section, we propose a novel methodology to incorporate dose–volume constraints that preserves the linearity of the formulation.

2.1.2. A linear programming formulation with dose–volume constraints. We may incorporate bounds on the mean dose received by all voxels in a given structure without destroying linearity of our model as follows:

$$\underline{M}_s \leqslant \frac{1}{|V_s|} \sum_{j \in V_s} D_j(\vec{u}) \leqslant \overline{M}_s \qquad s \in S. \tag{11}$$

Mean-dose constraints are often used in lieu of dose–volume constraints to control the sparing of critical structures. It is often deemed desirable, however, to impose additional constraints on the distribution of dose in a structure (Carol *et al* 1997). A measure commonly used by clinicians to judge the quality of a treatment plan is the cumulative dose–volume histogram (DVH). Using the finite voxel representation of structures, the fractional (as opposed to percentage) cumulative DVH for structure s under radiation intensities $\vec{u}$ can be expressed as

$$H_s(\delta; \vec{u}) = \frac{|\{j \in V_s : D_j(\vec{u}) \geqslant \delta\}|}{|V_s|} \qquad \delta \in \mathbb{R}^+ \tag{12}$$

where δ denotes a dose threshold. Typical dose–volume constraints limit the relative volume of a structure that receives more or less dose than a particular threshold. For example, an upper dose–volume constraint on a salivary gland might state that no more than 50% of its voxels may receive a dose exceeding 30 Gy. Similarly, a lower dose–volume constraint on a target might state that no more than 5% of its voxels may receive a dose that falls short of the prescription dose. An equivalent way of stating these constraints is that upper dose–volume constraints limit the minimum value of the upper tail of the differential DVH for a given upper fractional volume, and lower dose–volume constraints limit the maximum value of the lower tail for a given lower fractional volume. Restating the examples above in this equivalent manner, it is required that the minimum dose received by the 50% of the voxels in the salivary gland receiving the highest doses should not exceed 30 Gy, and the maximum dose received by the 5% of the voxels in the target receiving the lowest doses should not fall short of the prescription dose.

We cannot apply such traditional dose–volume constraints in a linear model without introducing integer-valued decision variables, leading to a MILP (Langer *et al* 1990). As we

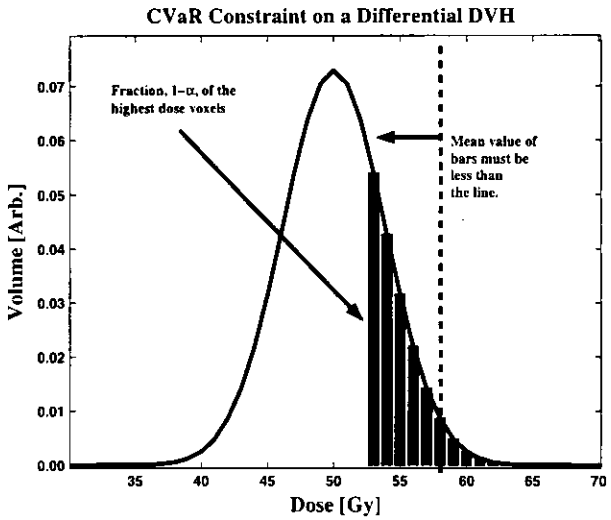


Figure 2. Schematic demonstrating the definition of a CVaR constraint.

pointed out in the introduction, the major disadvantage of MILP is that the typical problem size of clinically relevant problems prohibits finding an optimal solution to such models in an acceptable amount of computational time. Although we believe that imposing some type of constraints on the DVH is desirable since it allows for the division of voxels in a structure into two subsets that are treated differently, we believe that there is no fundamental basis for imposing the traditional type of dose–volume constraints other than the fact that many clinicians use the DVH to evaluate treatment plans. We propose to impose a novel alternative to the traditional dose–volume constraints described above that can be modelled within the framework of LP. To achieve this, we formulate conditional value-at-risk (CVaR) constraints on the differential DVH that, instead of bounding the minimum or maximum of a tail of the differential DVH, bound the mean value of these tails. More formally, our constraints are of the following form, for the lower and upper tails of the differential DVH, respectively:

(i) The average dose received by the subset of a target of relative volume $1-\alpha$ receiving the lowest doses must be at least equal to L^{α}.
(ii) The average dose received by the subset of a structure of relative volume $1-\alpha'$ receiving the highest doses may be no more than $U^{\alpha'}$.

For the sake of clarity, figure 2 illustrates an example of an upper CVaR constraint applied to a particular differential DVH. Note that CVaR constraints generalize the mean-dose constraints described above by constraining the mean dose received by subsets of voxels receiving the highest or lowest doses among all voxels in the structures rather than the mean of the entire dose distribution.

The motivation for our constraints on the differential DVH stems from the fact that the problem of controlling the shape of the differential DVH resembles, to a large extent, a problem that has received much attention in the recent financial engineering literature. This approach, which was originally developed by Rockafellar and Uryasev (2002), formulates risk management constraints in terms of so-called tail means of distributions of financial risk.

In the following, we will derive an expression for the tail mean dose, in the financial literature referred to as the conditional value-at-risk (CVaR), which is bounded in our

dose–volume constraints. In order to define the CVaR, however, we first need to discuss the concept of value-at-risk (VaR). The upper VaR at level α (sometimes called the upper α-VaR) is the smallest dose level with the property that no more than $100(1-\alpha)\%$ of a structure receives a larger dose. More formally, the upper α-VaR for structure s is defined as

$$\overline{\zeta}_s^{\alpha} = \min\{\zeta \in \mathbb{R} : H_s(\zeta; \vec{u}) \leqslant 1-\alpha\}. \tag{13}$$

The upper CVaR at level α (sometimes called the upper α-CVaR) is then the mean of all doses that exceed the upper α-VaR. More formally (and appropriately accounting for the case where $H_s(\overline{\zeta}_s^{\alpha}; \vec{u}) < 1-\alpha$), the upper α-CVaR is defined as

$$\overline{\phi}_s^{\alpha}(\vec{u}) = \overline{\zeta}_s^{\alpha}(\vec{u}) + \frac{1}{(1-\alpha)|V_s|}\sum_{j\in V_s}\max\left(0, D_j(\vec{u}) - \overline{\zeta}_s^{\alpha}(\vec{u})\right). \tag{14}$$

Similarly, the lower α-VaR is the largest dose level with the property that no more than $100(1-\alpha)\%$ of a structure receives a smaller dose. More formally, the lower α-VaR for structure s is defined as

$$\underline{\zeta}_s^{\alpha} = \max\{\zeta \in \mathbb{R} : H_s(\zeta; \vec{u}) \geqslant 1-\alpha\}. \tag{15}$$

The lower α-CVaR is then the mean of all doses that fall short of the lower α-VaR. More formally (and appropriately accounting for the case where $H_s(\underline{\zeta}_s^{\alpha}; \vec{u}) > 1-\alpha$), the lower α-CVaR is defined as

$$\underline{\phi}_s^{\alpha}(\vec{u}) = \underline{\zeta}_s^{\alpha}(\vec{u}) - \frac{1}{(1-\alpha)|V_s|}\sum_{j\in V_s}\max\left(0, \underline{\zeta}_s^{\alpha}(\vec{u}) - D_j(\vec{u})\right). \tag{16}$$

For each structure, we may now define a set of upper and lower CVaR bounds. For targets, we will impose lower bounds L_s^{α} for $\alpha \in \underline{A}_s$, where $\underline{A}_s$ is a (finite) subset of $(0, 1)$ ($s \in S^1$), and for all structures we will impose upper bounds U_s^{α} for $\alpha \in \overline{A}_s$, where $\overline{A}_s$ is a (finite) subset of $(0, 1)$ ($s \in S$). Summarizing, the CVaR constraints may be formulated as follows:

$$\underline{\phi}_s^{\alpha}(\vec{u}) \geqslant L_s^{\alpha} \qquad \alpha \in \underline{A}_s \quad s \in S^1 \tag{17}$$

$$\overline{\phi}_s^{\alpha}(\vec{u}) \leqslant U_s^{\alpha} \qquad \alpha \in \overline{A}_s \quad s \in S. \tag{18}$$

We next use the reformulation of constraints (17), (18) as described in Rockafellar and Uryasev (2002). This reformulation uses artificial variables $\underline{w}_j^{\alpha}$, $\underline{\zeta}_s^{\alpha}$, $\overline{w}_j^{\alpha}$ and $\overline{\zeta}_s^{\alpha}$ to achieve linearity; such use of artificial variables is a standard practice in the formulation of LP models.

Our complete LP model, including the mean-dose constraints and the reformulated CVaR constraints (which we will refer to as the LP problem with PWL objective function and CVaR constraints: $\mathrm{LP}_{\mathrm{PWL+CVaR}}$), now reads

$$\text{minimize} \sum_{j\in V} t_j \tag{19}$$

subject to

$$\Delta_j = \sum_{i\in N} D_{ij}u_i \qquad j \in V \tag{20}$$

$$a_k^{s_j}\Delta_j + b_k^{s_j} \leqslant t_j \qquad k = 1, \ldots, m_{s_j} \quad j \in V \tag{21}$$

$$L_{s_j} \leqslant \Delta_j \leqslant U_{s_j} \qquad j \in V \tag{22}$$

$$\underline{M}_s \leqslant \frac{1}{|V_s|}\sum_{j\in V_s}\Delta_j \leqslant \overline{M}_s \qquad s \in S \tag{23}$$

$$\underline{\zeta}_s^{\alpha} + \frac{1}{(1-\alpha)|V_s|} \sum_{j \in V_s} \underline{w}_j^{\alpha} \geqslant L_s^{\alpha} \qquad \alpha \in \underline{A}_s \quad s \in S^1 \tag{24}$$

$$\underline{w}_j^{\alpha} \geqslant \underline{\zeta}_s^{\alpha} - \Delta_j \qquad j \in V_s \quad \alpha \in \underline{A}_s \quad s \in S^1 \tag{25}$$

$$\overline{\zeta}_s^{\alpha} - \frac{1}{(1-\alpha)|V_s|} \sum_{j \in V_s} \overline{w}_j^{\alpha} \leqslant U_s^{\alpha} \qquad \alpha \in \overline{A}_s \quad s \in S \tag{26}$$

$$\overline{w}_j^{\alpha} \leqslant \Delta_j - \overline{\zeta}_s^{\alpha} \qquad j \in V_s \quad \alpha \in \overline{A}_s \quad s \in S \tag{27}$$

$$u_i \geqslant 0 \qquad i \in N \tag{28}$$

$$t_j, \Delta_j \geqslant 0 \qquad j \in V \tag{29}$$

$$\underline{w}_j^{\alpha} \geqslant 0 \qquad j \in V_s \quad \alpha \in \underline{A}_s \quad s \in S^1 \tag{30}$$

$$\overline{w}_j^{\alpha} \geqslant 0 \qquad j \in V_s \quad \alpha \in \overline{A}_s \quad s \in S \tag{31}$$

$$\underline{\zeta}_s^{\alpha} \text{ free} \qquad \alpha \in \underline{A}_s \quad s \in S^1 \tag{32}$$

$$\overline{\zeta}_s^{\alpha} \text{ free} \qquad \alpha \in \overline{A}_s \quad s \in S. \tag{33}$$

Unfortunately, there is in general no clear intuitive interpretation or practical significance of the values of the artificial decision variables $\underline{w}_j^{\alpha}$, $\underline{\zeta}_s^{\alpha}$, $\overline{w}_j^{\alpha}$ and $\overline{\zeta}_s^{\alpha}$. The variable ζ_s^{α} is a bound on the corresponding VaR, and an associated value of w_j^{α} is equal to the shortfall (in case of a lower CVaR constraint) or surplus (in case of an upper CVaR constraint) of the value of ζ_s^{α} by the dose received by voxel j. In case a CVaR bound is tight (satisfied as an equality) in the optimal solution to a LP problem with the above feasible region, its value of ζ_s^{α} is in fact equal to the VaR corresponding to that constraint.

As for the lower and upper bound constraints, lower and upper CVaR constraints can also be conflicting. We may define PWL penalty functions corresponding to surpluses or shortfalls of the CVaR constraints, and incorporate these linearly into the optimization problem in a similar way as we have done for the PWL voxel dose penalty constraints.

In case a target has invaded a critical structure to which we wish to apply an upper CVaR constraint, we often choose to reformulate this CVaR constraint to only apply to the part of the critical structure outside the target, i.e. the set of voxels in the critical structure for which this structure is dominant. For example, an upper α-CVaR constraint on a critical structure $s \in S^2$ with voxel set V_s is then replaced by an upper α-CVaR constraint on the reduced voxel set $\{j \in V_s : s_j = s\}$, where

$$\alpha' = 1 - \frac{(1-\alpha)|V_s| - |\{j \in V_s : s_j \neq s\}|}{|\{j \in V_s : s_j = s\}|}. \tag{34}$$

Note that if the relative volume of the part of the structure that is also in a target is larger than $1-\alpha$, we will obtain $\alpha' > 1$ and the CVaR constraint cannot be satisfied by any solution to the model.

2.2. *Model implementation and testing*

To test our models, three-dimensional image-based treatment-planning data for head-and-neck cancer patients were generated on our in-house treatment-planning system: the University of Florida optimized radiation therapy (UFORT) treatment planning decision support system (TPDSS). The UFORT TPDSS was developed using a commercial technical programming language (Matlab, Mathworks Inc.). This system was designed to accept DICOM communications of treatment planning image, structure and plan data from a commercial treatment planning simulation software (VoxelQ, Philips Medical Systems) in our clinic. The system anonymized the patient data for research purposes and converted the data to an internal data format. Users followed four steps to execute a FMO for treatment planning: (1) the isocentre to use for dose calculation was identified; (2) the critical organ and target-structure names were associated with unique structures on a list of expected structures; (3) prescription doses for targets were defined; and (4) the number and angles of beams were specified. Margins for penumbra were automatically generated for the union of the targets in each case, and asymmetric secondary jaw settings were determined. The beam apertures were then discretized into 1×1 cm^2 beamlets. An isotropic $3 \times 3 \times 3$ mm^3 dose voxel grid was generated, and the centroid of each dose voxel was given a unique label and tested for membership in all of the structures and recorded. To compute dose, we employed a simple empirical beamlet model that was fitted to radiochromic film data of 1×1 cm^2 beamlets formed by secondary divergent jaw collimators and measured using a previously reported methodology (Dempsey *et al* 2000). Dose due to leakage through the collimators was subtracted from the tails of the small beam data used for the fit as this information makes the D_{ij} data unnecessarily dense and does not accurately represent the leakage produced by the actual sequenced delivery. An example of the accuracy obtainable with this model is presented in figure 3. Our method required between 0.7 and 1 s of computational time per beamlet on a 2.8 GHz Pentium 4 computer. The data generated by this method were written to a double-precision floating-point sparse matrix of D_{ij} values.

A computer program was developed in C++ to interface with an industrial LP solver (CPLEX 8, ILOG Inc.). This interface program reads in the model data from our TPDSS and prepares it in a format (Concert Technologies, ILOG) known to the LP solver. Then the model is solved using the solver's implementation of the barrier interior-point method (Wright 1997). Once the model is solved to optimality, the optimal intensity vector, $\vec{u}^*$, is written to a file for the UFORT system. The optimal intensities were discretized for each beam angle to a user selectable percentage (in this case 5% levels) in preparation for leaf sequencing. The resulting plan dose distribution and histograms were computed by summing the D_{ij} weighted by the discretized intensities as in equation (1). Leaf-transmission leakage intensities were estimated at 1.7% for otherwise zero intensity bixels. The plans were then reviewed using a graphic user interface that allows exploration of structure, DVH and dose data.

We next discuss the form of the convex voxel-based penalty functions $F_s(\Delta_j)$ that we approximated for this example. For each structure $s \in S$, we specify a lower bound L_s and an upper bound U_s (where, for critical structures $s \in S^2$, we have $L_s = 0$), deviations from which are heavily penalized. In addition, each structure also has a threshold dose T_s, and deviations from this value are penalized using an asymmetric piecewise polynomial function:

$$F_s(\Delta_j) = \begin{cases} \beta_s(T_s - L_s)^{n_s} + \overline{\beta}_s(L_s - \Delta_j) & \text{if} \quad \Delta_j < L_s \\ \beta_s \max(T_s - \Delta_j, 0)^{n_s} + \gamma_s \max(\Delta_j - T_s, 0)^{m_s} & \text{if} \quad L_s \leqslant \Delta_j \leqslant U_s \\ \gamma_s(U_s - T_s)^{m_s} + \overline{\gamma}_s(\Delta_j - L_s) & \text{if} \quad \Delta_j > U_s \end{cases} \tag{35}$$

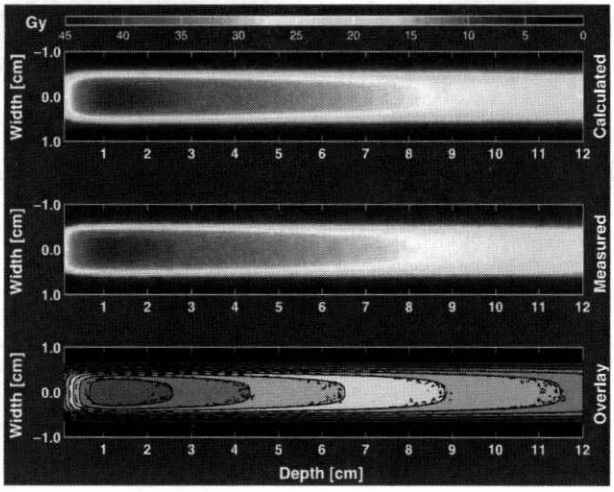


Figure 3. Examples of a $D_{max} = 45$ Gy beam with dose calculated (labelled 'calculated') for a 1×1 cm^2 beam in water using a simple pencil beam model fit to RCF data at 2×2 cm^2 and 5×5 cm^2, an RCF measurement (labelled 'measured') for a 1×1 cm^2 beam and an isodose curve overlay for the calculated (solid lines) and the measured (dashed lines) distributions with curves at 40, 35, 30, 25, 20, 15, 10, 5 and 2 Gy. Both the measurement and the calculation were performed for a 0.1 mm sized voxel.

where $\beta_s, \overline{\beta}_s, \gamma_s$ and $\overline{\gamma_s}$ are structure-dependent coefficients, and n_s and m_s are powers of the piecewise polynomial penalty function. The parameters β_s and n_s are associated with underdosing a voxel, while the parameters γ_s and m_s are associated with overdosing a voxel. To ensure convexity, note that we should choose $\beta_s, \gamma_s \geqslant 0$ and $n_s, m_s \geqslant 1$. The coefficients penalize deviations from the lower and upper bounds, and should be chosen large enough to ensure convexity of the penalty function. Note that, for critical structures $s \in S^2$ we always choose $L_s = \beta_s = 0$.

When CVaR constraints are added, we also allow deviations of the corresponding bounds at a very high penalty. To achieve this, we modify the CVaR lower and upper bound constraints by introducing additional artificial variables $\underline{\phi}_s^{\alpha}$ and $\overline{\phi}_s^{\alpha}$ as follows:

$$\underline{\zeta}_s^{\alpha} + \frac{1}{(1-\alpha)|V_s|} \sum_{j \in V_s} \underline{w}_j^{\alpha} = \underline{\phi}_s^{\alpha} \qquad \alpha \in \underline{A}_s \quad s \in S^1 \tag{36}$$

$$\overline{\zeta}_s^{\alpha} - \frac{1}{(1-\alpha)|V_s|} \sum_{j \in V_s} \overline{w}_j^{\alpha} = \overline{\phi}_s^{\alpha} \qquad \alpha \in \overline{A}_s \quad s \in S \tag{37}$$

$$\underline{\phi}_s^{\alpha} \text{ free} \qquad \alpha \in \underline{A}_s \quad s \in S^1 \tag{38}$$

$$\overline{\phi}_s^{\alpha} \text{ free} \qquad \alpha \in \overline{A}_s \quad s \in S. \tag{39}$$

We then add the following terms to the objective function:

$$\sum_{s \in S^1} \sum_{\alpha \in \underline{A}_s} \underline{G}_s^{\alpha}\left(\underline{\phi}_s^{\alpha}\right) + \sum_{s \in S} \sum_{\alpha \in \overline{A}_s} \overline{G}_s^{\alpha}\left(\overline{\phi}_s^{\alpha}\right) \tag{40}$$

Table 1. Values of the coefficients of the voxel-based penalty functions that were used in solving the illustrated case.

Structure (s)	T_s	L_s	$\underline{\beta}_s$	n_s	$\overline{\beta}_s$	U_s	$\overline{\gamma}_s$	m_s	$\overline{\gamma}_s$
PTV1	72.5	69.5	20	12	10^{11}	75.5	20	6	10^{10}
PTV2	52	49.5	7	12	10^{9}	55.5	7	6	10^{8}
Right parotid	0	0	0	–	–	75.5	500	4	10^{11}
Right submandibular	0	0	0	–	–	75.5	5500	4	10^{11}
Tissue	28	0	0	–	–	75.5	300	5	10^{11}
Spinal cord	40.5	0	0	–	–	45	0.5	2	10^{10}
Brainstem	45	0	0	–	–	50	0.6	2	10^{10}
Mandible	70	0	0	–	–	77	0.3	2	10^{6}

where

$$\underline{G}_s^{\alpha}\left(\underline{\phi}_s^{\alpha}\right) = \underline{\eta}_s^{\alpha} \max\left(0, L_s^{\alpha} - \underline{\phi}_s^{\alpha}\right) \tag{41}$$

$$\overline{G}_s^{\alpha}\left(\overline{\phi}_s^{\alpha}\right) = \overline{\eta}_s^{\alpha} \max\left(0, \underline{\phi}_s^{\alpha} - U_s^{\alpha}\right) \tag{42}$$

and $\underline{\eta}_s^{\alpha}, \overline{\eta}_s^{\alpha}$ are nonnegative slope parameters. Similar to the linear reformulation of the voxel-based penalty functions, we can use artificial variables to incorporate these additional terms into our LP model.

Finally, for defining a dominant structure for each voxel we have used a priority list in which targets have the highest priorities, followed by the critical structures, with the least important structure being unspecified tissue or skin. We note that when we relax the bound constraints and include a penalty for violating them, the use of dominant structures is not required, since any inconsistency can be sorted out by the optimization. However, we have chosen not to take this approach.

3. Results

The clinical goals of our optimization were to simultaneously cover 95% or more of two planning target volumes, PTV1 and PTV2, to doses of 70 and 50 Gy, with the minimum dose bounded within 7% of the prescription dose for both targets, and the maximum dose bounded within 10% of the prescription dose for PTV1 only. It was necessary to allow hot spots in PTV2 to obtain adequate coverage of PTV1, as PTV1 is a subset of PTV2. We also had the following requirements: at least one of four salivary glands should have 50% or less of its volume covered by 30 Gy or higher (Chao *et al* 2000); the spinal cord should have 99% or more of its volume covered by less than 45 Gy; the brainstem should have 99% of its volume covered by less than 50 Gy; the unspecified tissue (often referred to as 'skin' or 'tissue') should have 97% of its volume covered by less than 50 Gy.

A single case was first examined with seven equispaced beams having International Electrotechnical Commission (IEC) gantry angles of 0°, 51°, 103°, 154°, 206°, 257° and 309°. The UFORT system generated 1182 beamlets to adequately cover the targets from the seven beam angles, and the 3 mm isotropic voxel grid resulted in 206 152 voxels and generated 1 876 965 nonzero D_{ij} values in a sparse matrix of size 1182 by 206 152 (density: 0.77%) that were output by the planning system.

The parameters of the LP_{PWL} model were determined by manual adjustment and are shown in table 1. Similar to Tsien *et al* (2003), we found that high powers of dose difference

Table 2. Values of the coefficients corresponding to the CVaR constraints that were used in solving the illustrated case.

	Lower CVaR-constraints			Upper CVaR-constraints		
Structure (s)	α	L_s^α	$\underline{\eta}_s^\alpha$	α	U_s^α	$\bar{\eta}_s^\alpha$
PTV1	0.90	68	10^{12}	0.99	75	10^{11}
PTV2	0.95	48	10^{12}	0.95	55	10^{11}
Right parotid	–	–	–	0.60	26	10^{11}
Right submandibular	–	–	–	0.60	26	10^{11}

lead to excellent results. We chose to approximate the piecewise polynomial penalty functions for the targets by two PWL segments for underdosing and four segments for overdosing, and for the critical structures by three segments for overdosing (not counting the segments penalizing violations of the bounds). Ipsilateral (left) salivary glands were not spared due to their proximity to PTV1.

To demonstrate the utility of the CVaR constraints, we formulated an instance of the $LP_{PWL+CVaR}$ model that penalized both contralateral (right) salivary glands as tissue (i.e. we did not attempt to spare the glands using the objective function), while adding upper CVaR constraints to achieve sparing of these structures. In addition, we added lower and upper CVaR constraints on both targets. The corresponding coefficients were again determined by manual adjustment and are shown in table 2.

After preprocessing by the CPLEX solver, the model LP_{PWL} contained 538 334 constraints, 661 490 variables and 2 920 435 nonzero elements in the constraint matrix. The time needed to find the globally optimal solution was 302.5 s on a 2.8 GHz Pentium 4 laptop computer with 2 GB of RAM. The model $LP_{PWL+CVaR}$ contained 562 694 constraints, 685 850 variables and 3 177 467 nonzero elements in the constraint matrix. The time needed to find the globally optimal solution was 425 s.

Figure 4 shows cumulative DVHs for targets, spared salivary glands and tissue for both solutions. We found that, in both cases, nearly all of our planning goals were satisfied. In the solution to the LP_{PWL} model, target coverage was >95% for both target volumes at their prescription doses, with 100% coverage at 69 Gy for PTV1 and 99.6% coverage at 46.5 Gy for PTV2. The maximum dose to PTV1 was 78 Gy. Both contralateral salivary glands were spared, with 1% of the right parotid and 37.8% of the right submandibular gland receiving 30 Gy or higher. The spinal cord received a maximum dose of 47 Gy, with 1.2% exceeding 45 Gy. The maximum dose in the brainstem was 32 Gy. The unspecified tissue had 1.4% of its volume exceeding 50 Gy, and less than 0.1% of its volume exceeding 57 Gy. The mandible had 1% of its volume exceeding 70 Gy, and a maximum dose of 77 Gy. In the solution to the $LP_{PWL+CVaR}$ model, target coverage was slightly better, with 100% coverage at 69 Gy for PTV1, and 99.8% coverage at 46.5 Gy for PTV2. The maximum dose to PTV1 was 77 Gy. Both contralateral salivary glands were spared, with 3% of the right parotid and 35.8% of the right submandibular gland receiving 30 Gy or higher. The mean doses to these glands were 18 Gy and 28 Gy, respectively. The spinal cord received a maximum dose of 45 Gy. The maximum dose in the brainstem was 34 Gy. The unspecified tissue had 1.7% of its volume exceeding 50 Gy, and less than 0.1% of its volume exceeding 57 Gy. The mandible had 0.7% exceeding 70 Gy, and a maximum dose of 76 Gy.

Figure 5 illustrates the nearly equivalent salivary gland sparing obtained using the PWL voxel-based objective function or our CVaR dose–volume constraints. Both contralateral salivary glands were spared with significantly less than our limit of 50% at 30 Gy or higher.

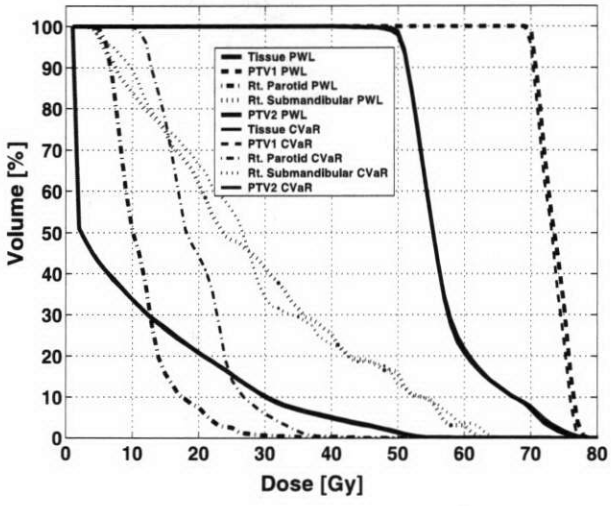


Figure 4. Cumulative DVHs of tissue (solid line), PTV1 (dashed line), right parotid (dash-dot line), right submandibular (dotted line), PTV2 (solid line), from the solutions of the FMO problem employing the LP_{PWL} (thick lines) and $LP_{PWL+CVaR}$ (thin lines) models.

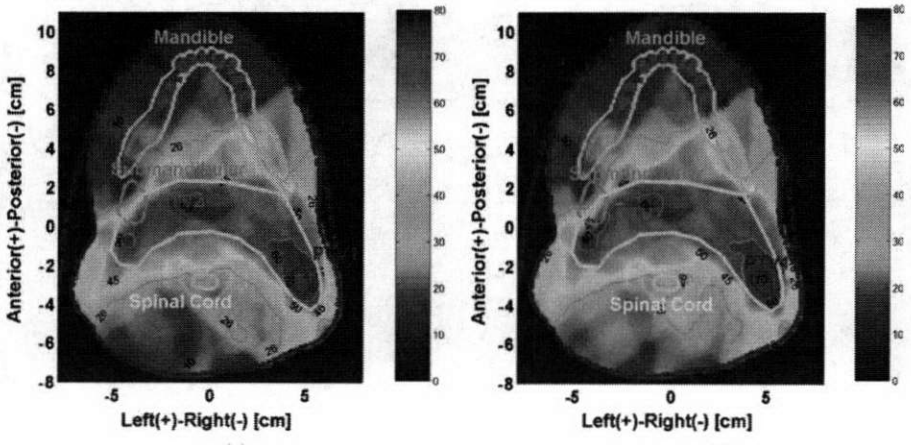


Figure 5. Overlay of axial CT isodose curves (10, 26, 45, 50, 60, 70 Gy), dose colourwash, and structures (PTV1-blue, PTV2-orange, spinal cord-yellow, right submandibular gland-green, mandible-cyan) for the globally optimal solutions obtained by (a) the LP_{PWL} model and (b) the $LP_{PWL+CVaR}$ model.

Noting that approximately 80% of the voxels in our model are in unspecified tissue of the case, we investigated the use of a coarser dose grid for this structure. Lowering the dose grid resolution from a 3 mm isotropic voxel grid to a 6 mm isotropic dose grid for only the unspecified tissue structure we found that, after preprocessing by the CPLEX solver, the LP_{PWL} model with the reduced unspecified tissue resolution contained 30 201 constraints,

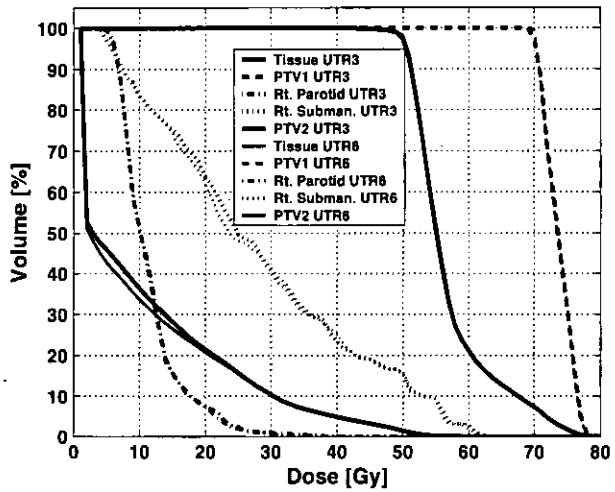


Figure 6. Cumulative DVHs of tissue (solid line), PTV1 (dashed line), right parotid (dash-dot line), right submandibular (dotted line), PTV2 (solid line), from the solutions of the FMO problem employing the LP_{PWL} with an unspecified tissue resolution of 3 mm (UTR3; thick lines) and with an unspecified tissue resolution of 6 mm (UTR6; thin lines).

166 514 variables and 701 496 nonzero elements in the constraint matrix. The time needed to find the globally optimal solution was 52.9 s on a 2.8 GHz Pentium 4 computer with 1 GB of RAM. The $LP_{PWL+CVaR}$ model with the reduced unspecified tissue resolution contained 190 874 constraints, 221 075 variables and 1 125 042 nonzero elements in the constraint matrix. The time needed to find the globally optimal solution was 125.8 s. Figure 6 demonstrates that the solutions of these models are nearly identical for the LP_{PWL} model. Similar results were found for the $LP_{PWL+CVaR}$ model.

We next investigated the influence of the number of segments used to approximate the nonlinear objective functions in both models by doubling and quadrupling the number of segments while using the reduced unspecified tissue resolution. Increasing the number of segments does not increase the number of constraints in the model. Doubling increased the number of variables in the LP_{PWL} model to 291 140, and the number of nonzero elements in the constraint matrix to 826 122, and the computation time to 65 s. Quadrupling increased the number of variables to 540 392, and the number of nonzero elements in the constraint matrix to 1 075 374, and the computation time to 96.3 s. The computation time increases approximately linearly with the number of segments in the model. Figure 7 demonstrates that there is essentially no difference between the obtained solutions using the LP_{PWL} model. A small difference was observed in the DVHs for unspecified tissue at lower doses. This difference is insignificant and would not change any clinical decisions. Similar results were found for the $LP_{PWL+CVaR}$ model.

Finally, we investigated the robustness of the parameters that were obtained for the single case discussed above using manual adjustment by applying our model to seven additional head-and-neck IMRT cases where definitive therapy and salivary gland sparing was desired. We again used the reduced unspecified tissue voxel grid. The objective function parameters in tables 1 and 2 were scaled by the ratios of the number of voxels in corresponding structures, to ensure that the relative importance of each structure remains the same in each case. The

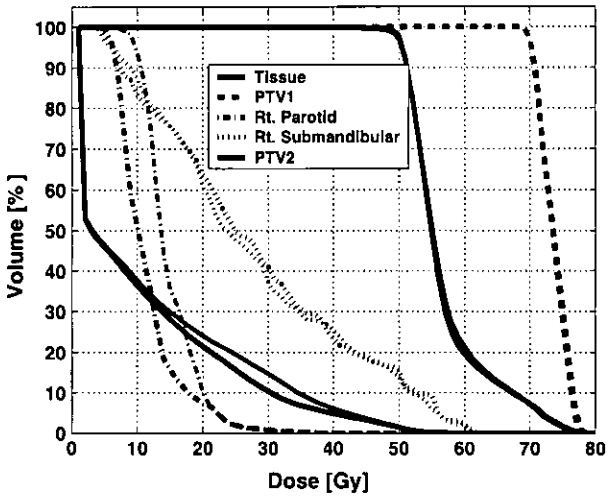


Figure 7. Overlay of three sets of DVHs for the original segments (thick lines), doubled number of segments (medium lines) and quadrupled number of segments (thin lines).

Table 3. Model size and run times.

Case	Model	Run time (s)	Constraints	Variables	Number of nonzero elements	Number of voxels	Number of beamlets
1	PWL	52.9	30 201	166 514	701 496	40 984	1 182
	CVaR	125.8	190 874	221 075	1 125 042		
2	PWL	127.0	63 127	358 125	1 427 730	82 540	1 745
	CVaR	267.9	369 129	432 345	1 880 735		
3	PWL	151.3	66 205	368 824	1 474 210	90 967	1 804
	CVaR	434.8	426 072	492 277	2 440 754		
4	PWL	36.8	29 871	165 898	696 882	41 523	1 017
	CVaR	88.7	173 856	203 727	923 095		
5	PWL	40.0	26 156	156 448	655 604	32 942	1 015
	CVaR	84.9	30 007	161 580	694 098		
6	PWL	55.5	38 831	231 925	936 477	52 800	1 178
	CVaR	132.2	240 007	278 840	1 241 489		
7	PWL	54.3	35 121	212 958	876 606	45 688	1 090
	CVaR	87.8	39 325	218 133	920 971		
8	PWL	37.2	26 173	155 868	644 395	36 566	921
	CVaR	84.5	161 012	187 185	846 979		

coefficients for the left parotid and submandibular glands were chosen equal to the ones for the right parotid and submandibular glands. We only attempted to spare a salivary gland if it was sufficiently far away from PTV1 (>1 cm) and more than 50% of its volume was outside PTV2. The sizes and run times of these models are found in table 3. Note that the number of constraints scales with the number of voxels, which explains the reduction in model size when the reduced voxel grid for unspecified tissue is used. The number of variables scales with both the number of voxels and the number of segments used to approximate the PWL penalty functions. The running times of LP_{PWL} ranged between 37 and 151 s, and applying the CVaR

 3369

Table 4. UFORT automated treatment-planning results for eight head-and-neck cases using LP_{PWL}.

Structure	Dose–volume result	1	2	3	4	5	6	7	8
All	Pass/Fail	Pass	Pass	Pass	Pass	*Fail*	Pass	Pass	Pass
PTV1	$\% \geqslant D_{R_x}$	95	95	95	95	95	95	95	95
PTV1	$\% \geqslant .93D_{R_x}$	100	100	100	100	100	100	100	100
PTV1	$\% \geqslant 1.1D_{R_x}$	0.5	0	1.2	0	0.9	1.5	0.2	0
PTV2	$\% \geqslant D_{R_x}$	97.2	98.6	98.0	97.1	97.3	98.1	97.6	93.5
PTV2	$\% \geqslant .93D_{R_x}$	99.6	99.9	99.6	99.4	99.6	99.5	99.4	98.9
Left parotid	$\% \geqslant 30$ Gy	–	27.1	7.4	35.2	*56.5*	9.9	47.3	7.0
	Mean dose		15.8	11.2	27.2	*36.8*	16.3	29.3	14.8
Right parotid	$\% \geqslant 30$ Gy	1.0	19.4	16.0	40.4	*63.2*	–	21.4	*71.1*
	Mean dose	10.4	12.9	15.7	27.8	*34.9*		16.9	39.9
Left submandibular	$\% \geqslant 30$ Gy	–	–	27.8	–	–	42.4	–	44.8
	Mean dose			24.1			26.2		28.7
Right submandibular	$\% \geqslant 30$ Gy	37.8	–	*70.8*	–	–	–	*76.9*	–
	Mean dose	27.0		*36.8*				*47.3*	
Spinal cord	$\% \geqslant 45$ Gy	1.1	0	1.3	0	0	0	0	0
Brainstem	$\% \geqslant 54$ Gy	0	–	0	0	*1.9*	0	0	0
Tissue	$\% \geqslant 50$ Gy	1.4	1.8	1.5	1.2	1.1	1.3	1.5	1.0

Table 5. UFORT automated treatment-planning results for eight head-and-neck cases using $LP_{PWL+CVaR}$.

Structure	Dose–volume result	1	2	3	4	5	6	7	8
All	Pass/Fail	Pass	Pass	Pass	Pass	*Fail*	Pass	Pass	Pass
PTV1	$\% \geqslant D_{R_x}$	95	95	95	95	95	95	95	95
PTV1	$\% \geqslant .93D_{R_x}$	100	100	100	100	100	100	100	100
PTV1	$\% \geqslant 1.1D_{R_x}$	1.8	0	0	0	0	1.3	0	0.7
PTV2	$\% \geqslant D_{R_x}$	98.9	96.1	98.5	96.4	96.6	97.4	98	96.7
PTV2	$\% \geqslant .93D_{R_x}$	99.8	99.9	99.8	98.9	99.1	99.3	99.5	99.2
Left parotid	$\% \geqslant 30$ Gy	–	25.6	14.2	18.9	31.4	11.8	29.8	9.9
	Mean dose		15.5	17.7	21.2	27.2	17.2	25.0	20.9
Right parotid	$\% \geqslant 30$ Gy	3.0	20.8	17.4	24.1	37.5	–	18.6	*72.4*
	Mean dose	18.0	14.5	18.8	22.6	27.4		23.0	*39.8*
Left submandibular	$\% \geqslant 30$ Gy	–	–	29.3	–	–	*60.0*	–	45.6
	Mean dose			23.9			*32.0*		30.6
Right submandibular	$\% \geqslant 30$ Gy	35.8	–	*93.8*	–	–	–	*95.0*	–
	Mean dose	28.0		*46.8*				*46.6*	
Spinal cord	$\% \geqslant 45$ Gy	0.2	0	0	0	0	0	0	0
Brainstem	$\% \geqslant 54$ Gy	0	–	0	0	*2.5*	0	0	0
Tissue	$\% \geqslant 50$ Gy	1.4	1.8	1.5	1.2	1.1	1.3	1.5	1.0

constraints increased the running time by a factor of 2–3. Tables 4 and 5 display dose–volume data to compare to our criteria for all plans obtained using the LP_{PWL} and $LP_{PWL+CVaR}$ models, respectively. In addition, mean-dose data for the salivary glands is provided for reference. In all tables, case 1 refers to the case described in more detail above. We found that the model parameters for LP_{PWL} determined for case 1 produced excellent plans for seven out of the eight cases. Plan 5 failed to spare any salivary glands and failed to adequately spare

the brainstem. Strictly speaking, plan 8 failed to adequately cover PTV2 at 50 Gy, although 99.4% of the prescription dose for PTV2 covered 95% of the target. In all cases other than case 5, all criteria were satisfied to within 2% of their values (values that exceed this 2% limit are italicized in tables 4 and 5). Note that the plans were run in an automated fashion without manual adjustment of the problem parameters. Table 5 shows that when the $LP_{PWL+CVaR}$ model was applied, sparing of both parotid glands was achieved for case 5. Although the brainstem in that plan was still not adequately spared, inspection of the 3D planning data revealed that PTV1 and PTV2 came as close as 3 and 1 mm, respectively, to the brainstem, creating a situation where a clinical tradeoff would have to be made by a human user. Adopting a strategy of first running the LP_{PWL} model, and then running the $LP_{PWL+CVaR}$ model only if not all criteria are satisfied would result in an efficient and effective methodology for arriving at plans of good clinical quality.

4. Discussion and conclusions

4.1. Discussion

Our initial implementation of the LP_{PWL} and the $LP_{PWL+CVaR}$ models has demonstrated the ability to produce high quality results. We hypothesize that the quality of the results obtained is, to a large extent, due to the fact that the algorithm employed finds globally optimal solutions. Further studies could test this hypothesis by comparing our method to local search methods. Recent studies have attempted to heuristically search the hyper-dimensional solution space (typically in the thousands of dimensions) of the FMO problem (Llacer *et al* 2003, Wu and Mohan 2002, Wu *et al* 2003). Their methods find multiple convergence points, establishing the existence of local minima. Although it is found that the qualities of many local minima are similar, local minima that differ significantly from the global minimum are also found. This means that the issue of non-globally optimal local minima is an important one, even if the solutions found with a particular method appear to be clinically acceptable. The LP_{PWL} and the $LP_{PWL+CVaR}$ models completely side-step the issue of local minima as they are unimodal, and our LP algorithm provides a very efficient global optimization technique. Our results demonstrate that this combination of a unimodal model formulation with an efficient algorithm that is capable of recognizing globally optimal solutions can produce results of high clinical quality. Further, we have also demonstrated that in many cases these solutions can be achieved without the use of dose–volume constraints. However, it appears that in some cases dose–volume constraints are needed to split a structure into two parts that are treated differently in the model. Finally, we believe that the use of a global optimization method will make our approach more robust than the techniques that have found common use in commercial systems.

Previous studies have indicated that the quality of the dose computational algorithm employed can influence the quality of FMO solutions in the head-and-neck on the order of 5% (Jeraj *et al* 2002). Thus we are pursuing improved methods of dose computation for the UFORT TPDSS including rapid Monte Carlo simulation (Sempau *et al* 2000) and convolution-based methods (Ahnesjo and Aspradakis 1999). However, the optimization formulation and algorithm are essentially dose model independent and such changes are not expected to change the performance of the approach. Despite the considerable size of the models, application of the barrier interior-point method allows for very efficient global solutions with run times of approximately 30–150 s on a modern single processor laptop computer for our model with piecewise linear penalty functions, and a factor of 2–3 more time for our model with dose–volume constraints. The speed of the algorithm can allow for the development of

more interactive systems where many global solutions are produced in a single session by altering the objective function and constraints to manually steer the solution to a desired result. Extensions of the models we are currently investigating include: the inclusion of additional beamlet variables and constraints to allow for fractionated treatments; the addition of spatial information to the problem, either by suitably modifying the objective function or by incorporating sequences of CVaR constraints applied to subsets of structures; and the incorporation of discrete beam-orientation optimization without integer variables (i.e. not a MILP model). We are also currently working on the complete automation of the process of solving the FMO problem for head-and-neck cancer treatment plans. We feel that this goal should be obtainable given our theoretically robust technique for solving the FMO problem. This would be best done by optimizing the model parameters using a large population of patient cases. Finally, while currently not efficient (or perhaps even practical due to our model size), it would also be very interesting to incorporate our model into the MILP and multicriteria optimization frameworks.

4.2. Conclusions

Our approach of approximating any convex objective function by a PWL convex approximation allows us to retain the flexibility offered by general convex objective functions, while formulating the FMO problem as a LP problem. By modelling the problem of controlling the shape of the differential DVH as a financial risk management problem (Rockafellar and Uryasev 2002), we have formulated a novel type of differential DVH constraint, called a CVaR constraint, that constrains the mean dose received by subsets of voxels receiving the highest or lowest doses among all voxels in a given structure. The LP_{PWL} and the $LP_{PWL+CVaR}$ models presented in this work have been demonstrated to be flexible, fast, and provide globally optimal solutions with high clinical quality. While it appears that the LP_{PWL} model alone can produce excellent results in most cases, our novel CVaR-based dose–volume constraints can provide more model flexibility to allow additional sparing. The speed of the algorithm can allow for the development of more interactive systems where many globally optimal solutions are produced in a single session by altering the objective function and constraints to manually steer the solution to a desired result.

References

Ahnesjo A and Aspradakis M M 1999 Dose calculations for external photon beams in radiotherapy *Phys. Med. Biol.* **44** R99–R155

Alber M, Meedt G, Nusslin F and Reemtsen R 2002 On the degeneracy of the IMRT optimization problem *Med. Phys.* **29** 2584–9

Alber M and Nusslin F 1999 An objective function for radiation treatment optimization based on local biological measures *Phys. Med. Biol.* **44** 479–93

Bahr G K, Kereiake J G, Horwitz H, Finney R, Galvin J and Goode K 1968 Method of linear programming applied to radiation treatment planning *Radiology* **91** 686–93

Bednarz G *et al* 2002 The use of mixed-integer programming for inverse treatment planning with pre-defined field segments *Phys. Med. Biol.* **47** 2235–45

Bortfeld T 1999 Optimized planning using physical objectives and constraints *Semin. Radiat. Oncol.* **9** 20–34

Carol M P, Nash R V, Campbell R C, Huber R and Sternick E 1997 The development of a clinically intuitive approach to inverse treatment planning: partial volume prescription and area cost function *Proc. 12th ICCR (Salt Lake City, Utah)* pp 317–9

Censor Y, Altschuler M D and Powlis W D 1988 On the use of Cimmino's simultaneous projections method for computing a solution of the inverse problem in radiation therapy treatment planning *Inverse Probl.* **4** 607–23

Chao K S C, Low D A, Perez C A and Purdy J A 2000 Intensity-modulated radiation therapy in head and neck cancers: the Mallinckrodt experience *Int. J. Cancer* **90** 92–103

Deasy J O 1997 Multiple local minima in radiotherapy optimization problems with dose–volume constraints *Med. Phys.* **24** 1157–61

Dempsey J F, Low D A, Mutic S, Kirov A S, Nussbaum G H and Williamson J F 2000 Validation of a precision radiochromic film dosimetry system for quantitative two-dimensional imaging of acute exposure dose distributions *Med. Phys.* **27** 2462–75

Fogliata A, Bolsi A and Cozzi L 2003 Comparative analysis of intensity modulation inverse planning modules of three commercial treatment planning systems applied to head and neck tumour model *Radioth. Oncol.* **66** 29–40

Hajek B 1988 Cooling schedules for optimal annealing *Math. Oper. Res.* **13** 311–29

Hamacher H W and Küfer K H 2002 Inverse radiation therapy planning—a multiple objective optimization approach *Discrete Appl. Math.* **118** 145–61

Hou Q, Wang J, Chen Y and Galvin J M 2003 An optimization algorithm for intensity modulated radiotherapy—the simulated dynamics with dose–volume constraints *Med. Phys.* **30** 61–8

Hou Q and Wang Y G 2001 Molecular dynamics used in radiation therapy *Phys. Rev. Lett.* **87** 168101

Jeraj R, Keall P J and Siebers J V 2002 The effect of dose calculation accuracy on inverse treatment planning *Phys. Med. Biol.* **47** 391–407

Jones L C and Hoban P W 2000 Treatment plan comparison using equivalent uniform biologically effective dose (EUBED) *Phys. Med. Biol.* **45** 159–70

Kallman P, Lind B K and Brahme A 1992 An algorithm for maximizing the probability of complication-free tumor-control in radiation-therapy *Phys. Med. Biol.* **37** 871–90

Langer M, Brown R, Urie M, Leong J, Stracher M and Shapiro J 1990 Large-scale optimization of beam weights under dose–volume restrictions *Int. J. Radiat. Oncol. Biol. Phys.* **18** 887–93

Langer M and Leong J 1987 Optimization of beam weights under dose–volume restrictions *Int. J. Radiat. Oncol. Biol. Phys.* **13** 1255–60

Langer M, Morrill S, Brown R, Lee O and Lane R 1996 A comparison of mixed integer programming and fast simulated annealing for optimizing beam weights in radiation therapy *Med. Phys.* **23** 957–64

Lee E K, Fox T and Crocker I 2003 Integer programming in intensity-modulated radiation therapy treatment planning *Ann. Oper. Res.* **119** 165–81

Llacer J, Deasy J O, Portfeld T R, Solberg T D and Promberger C 2003 Absence of multiple local minima effects in intensity modulated optimization with dose–volume constraints *Phys. Med. Biol.* **48** 183–210

Lof J 2000 Development of a general framework for optimization of radiation therapy *PhD Thesis* Stockholm University, Sweden

Mavroidis P, Lind B K and Brahme A 2001 Biologically effective uniform dose (*D*) for specification, report and comparison of dose response relations and treatment plans *Phys. Med. Biol.* **46** 2607–30

Morrill S M, Lane R G, Jacobson G and Rosen I I 1991a Treatment planning optimization using constrained simulated annealing *Phys. Med. Biol.* **36** 1341–61

Morrill S M, Lane R G, Wong J A and Rosen I I 1991b Dose–volume considerations with linear-programming optimization *Med. Phys.* **18** 1201–10

Murty K 1976 *Linear Programming and Combinatorial Optimization* (New York: Wiley)

Niemierko A 1997 Reporting and analyzing dose distributions: a concept of equivalent uniform dose *Med. Phys.* **24** 103–10

Niemierko A, Urie M and Goitein M 1992 Optimization of 3D radiation-therapy with both physical and biological end-points and constraints *Int. J. Radiat. Oncol. Biol. Phys.* **23** 99–108

Rockafellar R T and Uryasev S 2002 Conditional value-at-risk for general loss distributions *J. Banking Finance* **26** 1443–71

Romeijn H E and Smith R L 1994 Simulated annealing for constrained global optimization *J. Glob. Optim.* **5** 101–26

Rosen I I, Lane R G, Morrill S M and Belli J A 1991 Treatment plan optimization using linear programming *Med. Phys.* **18** 141–52

Sempau J, Wilderman S J and Bielajew A F 2000 DPM, a fast, accurate Monte Carlo code optimized for photon and electron radiotherapy treatment planning dose calculations *Phys. Med. Biol.* **45** 2263–91

Shepard D M, Ferris M C, Olivera G H and Mackie T R 1999 Optimizing the delivery of radiation therapy to cancer patients *SIAM Rev.* **41** 721–44

Shepard D M, Olivera G H, Reckwerdt P J and Mackie T R 2000 Iterative approaches to dose optimization in tomotherapy *Phys. Med. Biol.* **45** 69–90

Siddon R L 1985 Prism representation: a 3D ray-tracing algorithm for radiotherapy applications *Phys. Med. Biol.* **30** 817–24

Tsien C, Eisbruch A, McShan D, Kessler M, Marsh R and Fraass B 2003 Intensity-modulated radiation therapy (IMRT) for locally advanced paranasal sinus tumors: incorporating clinical decisions in the optimization process *Int. J. Radiat. Oncol. Biol. Phys.* **55** 776–84

Webb S 1994 Optimizing the planning of intensity-modulated radiotherapy *Phys. Med. Biol.* **39** 2229–46

Webb S 1995 Optimizing radiation-therapy inverse treatment planning using the simulated annealing technique *Int. J. Imaging Syst. Technol.* **6** 71–9

Webb S 2001 *Intensity-Modulated Radiation Therapy* (Bristol: Institute of Physics Publishing)

Webb S and Lomax T 2001 There is no IMRT? *Phys. Med. Biol.* **46** L7–L8

Wright S J 1997 *Primal-Dual Interior-Point Methods* (Philadelphia, PA: Society for Industrial & Applied Mathematics)

Wu C, Jeraj R and Mackie T R 2003 The method of intercepts in parameter space for the analysis of local minima caused by dose–volume constraints *Phys. Med. Biol.* **48** N149–N157

Wu Q and Mohan R 2002 Multiple local minima in IMRT optimization based on dose–volume criteria *Med. Phys.* **29** 1514–27

Wu Q W, Djajaputra D, Wu Y, Zhou J N, Liu H H and Mohan R 2003 Intensity-modulated radiotherapy optimization with gEUD-guided dose–volume objectives *Phys. Med. Biol.* **48** 279–91

Wu Q W, Mohan R, Niemierko A and Schmidt-Ullrich R 2002 Optimization of intensity-modulated radiotherapy plans based on the equivalent uniform dose *Int. J. Radiat. Oncol. Biol. Phys.* **52** 224–35

Xing L, Hamilton R J, Spelbring D, Pelizzari C A, Chen G T Y and Boyer A L 1998 Fast iterative algorithms for three-dimensional inverse treatment planning *Med. Phys.* **25** 1845–1849

Fast voxel and polygon ray-tracing algorithms in intensity modulated radiation therapy treatment planning

Christopher Fox
Department of Radiation Oncology, University of Florida, P. O. Box 100385, Gainesville, Florida 32610-0385

H. Edwin Romeijn
Department of Industrial and Systems Engineering, University of Florida, P. O. Box 116595, Gainesville, Florida 32611-6595

James F. Dempsey[a)]
Department of Radiation Oncology, University of Florida, P.O. Box 116595, Gainesville, Florida 32611-6595

(Received 19 June 2005; revised 28 February 2006; accepted for publication 2 March 2006; published 25 April 2006)

We present work on combining three algorithms to improve ray-tracing efficiency in radiation therapy dose computation. The three algorithms include: An improved point-in-polygon algorithm, incremental voxel ray tracing algorithm, and stereographic projection of beamlets for voxel truncation. The point-in-polygon and incremental voxel ray-tracing algorithms have been used in computer graphics and nuclear medicine applications while the stereographic projection algorithm was developed by our group. These algorithms demonstrate significant improvements over the current standard algorithms in peer reviewed literature, i.e., the polygon and voxel ray-tracing algorithms of Siddon for voxel classification (point-in-polygon testing) and dose computation, respectively, and radius testing for voxel truncation. The presented polygon ray-tracing technique was tested on 10 intensity modulated radiation therapy (IMRT) treatment planning cases that required the classification of between 0.58 and 2.0 million voxels on a 2.5 mm isotropic dose grid into 1–4 targets and 5–14 structures represented as extruded polygons (a.k.a. Siddon prisms). Incremental voxel ray tracing and voxel truncation employing virtual stereographic projection was tested on the same IMRT treatment planning cases where voxel dose was required for 230–2400 beamlets using a finite-size pencil-beam algorithm. Between a 100 and 360 fold cpu time improvement over Siddon's method was observed for the polygon ray-tracing algorithm to perform classification of voxels for target and structure membership. Between a 2.6 and 3.1 fold reduction in cpu time over current algorithms was found for the implementation of incremental ray tracing. Additionally, voxel truncation via stereographic projection was observed to be 11–25 times faster than the radial-testing beamlet extent approach and was further improved 1.7–2.0 fold through point-classification using the method of translation over the cross product technique. © *2006 American Association of Physicists in Medicine.* [DOI: 10.1118/1.2189712]

I. INTRODUCTION

Improvements in polygon and voxel ray-tracing algorithms have not been presented in the radiation-therapy literature since the seminal works of Siddon in the 1980's.[1,2] This investigation presents algorithms that significantly outperform the polygon[1] and voxel[2] ray-tracing algorithms of Siddon for implementation in voxel classification and dose computation, respectively. In adaptive[3,4] intensity modulated radiation therapy (IMRT), with image-guidance, rapid computation of dose for fluence map optimization (FMO)[5–7] is required in order to limit the risk of errors arising from patient motion. However, ray tracing algorithms are currently the time limiting factor in most beamlet dose models for FMO where the majority of time can be expended performing polygon and voxel based ray-tracing computations (a profiler showed ~70–80% of the time in our in-house treatment planning system is devoted to ray tracing) through the millions of voxels associated with the targets and critical structures in the patient. The current practice in state-of-the-art, IMRT treatment planning, is to perform these calculations only once before the commencement of therapy and adopt them for the treatment duration. However, in reality the patient geometry is dynamic and can deviate due to set-up errors, intra- and inter-fraction organ motion, and physiological changes in the patient. To account for changes during therapy, IMRT planning is not only required on a more frequent basis, but also on a much condensed time scale to limit further geometrical uncertainties arising while treatment planning computations are performed. Multiple volumetric images obtained throughout a course of treatment can significantly improve the ability to account for inter- and eventually intra-fraction patient motion. However, processing information provided by additional volumetric imaging studies can be time consuming when used to re-optimize treatment. The proposed first step for adaptive IMRT requires rapid voxel classification and dose computation for FMO such that the inter-fraction re-evaluation can occur without treatment

 0094-2405/2006/33(5)/1364/8/$23.00

 3575

interruption.[5–7] Beamlet dose computation models used in IMRT FMO rely heavily on ray-tracing algorithms for two reasons. Firstly, for voxel classification,[1] i.e., to classify voxel centers as inside or outside of segmented targets and critical structures and also to establish whether voxels are inside or outside the path of a radiation beam. Secondly, to compute the exact radiological path traversed through the patient[2] to correct for tissue heterogeneities in the dose model. As each beamlet is considered separately, the radiological path problem typically involves millions of beamlet-voxel combinations. Consequently, the need for more efficient ray-tracing algorithms will continue to exist in radiation therapy medical physics. To this end, ray-tracing algorithms that significantly improve the computational efficiency of IMRT treatment planning, when compared to the existing methods are introduced.

II. MATERIALS AND METHODS

A. A polygon translation ray-tracing technique for voxel classification

Radiation therapy targets and critical structures are generally represented as either collections of voxels or extruded three-dimensional (3D) polygonal objects (a.k.a. Siddon prisms) that delineate a collection of voxels. Voxel classification is the identification of the voxels that reside within given targets and/or critical structures, and hence are contained within 3D polygons, where the 3D polygon representations of targets and critical structures are defined using radiographic landmarks on a set of imaging voxel data. An independent isotropic voxel dose grid is typically sampled from the imaging voxel data for dose computation and 3D polygons are then used to classify the voxels, i.e., determine if their centers are inside or outside of the 3D polygons. Several efficient algorithms that address the point-in-polygon problem have been put forward in the computer graphics literature.[8–10] However, there are varying computational costs associated with each technique and care must be taken in choosing the algorithm best suited to the task. When dealing with polygons with large numbers of vertices as in radiation therapy and when no *a priori* information is available for preprocessing methods based on ray tracing,[9] such as the method tested by Siddon,[1] have been found to be the most efficient.

To determine if a point, say $\vec{Q}$, is inside of a polygon, ray tracing is performed between a point known to be outside the polygon, say $\vec{P}$, and $\vec{Q}$. The point $\vec{Q}$ is inside if and only if there are an odd number of ray-polygon intersections between $\vec{P}$ and $\vec{Q}$.[1] By convention, points along the ray on the polygon boundary are considered to be inside the polygon. An example of the 2D projection of a 3D polygon and a row of data from the isotropic dose grid is represented in Fig. 1. In Siddon's algorithm, each row of points is traced from $\vec{P}$ to $\vec{Q}$, splitting the plane of the polygon into three distinct regions, above below and on the ray. The dimension orthogonal to the 2D plane of the polygon can then be checked for point classification. The sign of the cross products of the

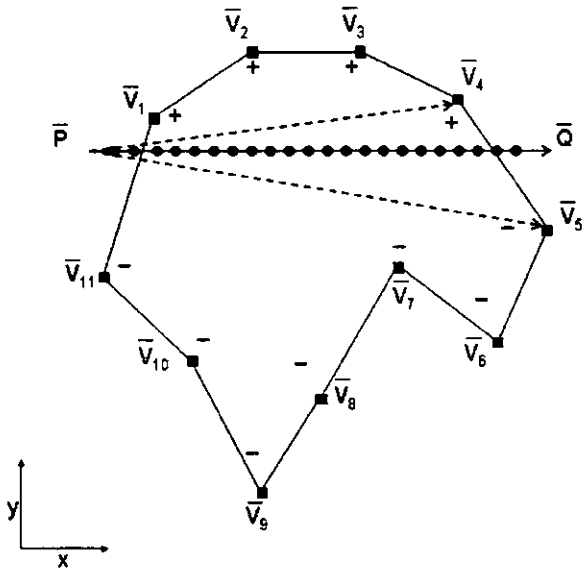


FIG. 1. Siddon's point-in-polygon technique demonstrates a 2D projection of a 3D polygon and a row of data from the isotropic dose grid. Rows of points are traced from $\vec{P}$ to $\vec{Q}$ splitting the plane into three distinct regions, above, below and on the ray. The region containing each vertex $\vec{V}_k$ is obtained from the cross-product sign of the rays $\overline{PQ}$ and $\overline{PV}$. A change in cross-product sign between consecutive vertices indicates an intersection of $\overline{PQ}$ and the polygon. Any point is inside the polygon if an odd number of intersections precede it.

tracing ray, $\overline{PQ}$, and the rays passing from $\vec{P}$ to the vertices $\overline{V_k}$ $(k=1,\ldots,n)$ of the polygon, $\overline{PV_k}$, indicate which region each vertex is in and is denoted by $\overline{PQ}\times\overline{PV_k}$. Observation of a cross-product sign change between consecutive vertices, say $\overline{V_k}$ and $\overline{V_{k+1}}$, is indicative of a ray-polygon intersection. Any point $\vec{T}$ along the ray $\overline{PQ}$ can be obtained from the parametric equation of a line given below:

$$\vec{T}=\vec{P}+\alpha(\vec{Q}-\vec{P}), \tag{1}$$

where α, called the intersection parameter, is the fractional length of $\overline{PQ}$ at which a crossing occurs and can be calculated as follows:

$$\alpha=\frac{\overline{PV_k}\times\overline{PV_{k+1}}}{\overline{PQ}\times\overline{PV_k}-\overline{PQ}\times\overline{PV_{k+1}}}. \tag{2}$$

The special case in which $\overline{PQ}\times\overline{PV_k}=0$ indicates an intersection of the tracing ray and the vertex and must be treated separately. The majority of the computational effort required by the Siddon polygon ray-tracing algorithm results from the repeated evaluation of cross products. This can lead to computation periods of tens of minutes for the large polygons encountered in radiation therapy treatment planning.

In our implementation, a bounding box is first obtained from the limits of the polygon vertices to restrict the size of the classification data set and identify a point $\vec{P}$ outside the polygon, i.e., a small finite distance outside the box as shown in Fig. 2. An algorithm based on the polygon translation technique of MacMartin, which is discussed by Nessar *et al.*[11] is then implemented. The polygon translation tech-

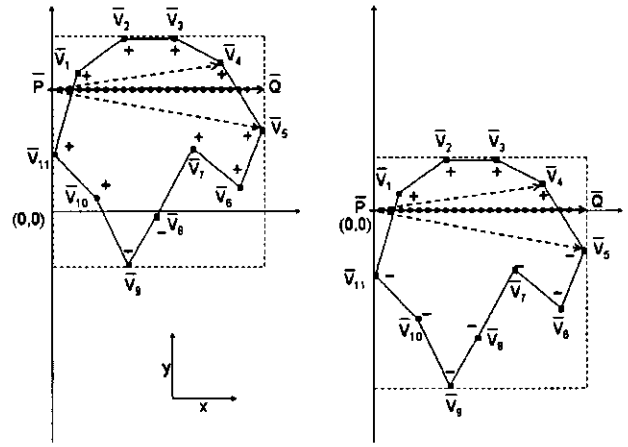


FIG. 2. Point-in-polygon translation technique. 2D projection of a 3D polygon and a row of data from the isotropic dose grid. A bounding box, set by the limits of the polygon, restricts the size of the data set. The coordinate system is translated so $\bar{P}$ lies at the origin. Rows of points are traced from $\bar{P}$ to $\bar{Q}$ splitting the plane into three distinct regions, above, below and on the ray. The region containing the vertex $\bar{V}$ is determined from the sign of the vertex Y coordinate. A change the sign of Y between consecutive vertices indicates an intersection of $\overline{PQ}$ and the polygon. Any point is inside the polygon if an odd number of intersections precede it.

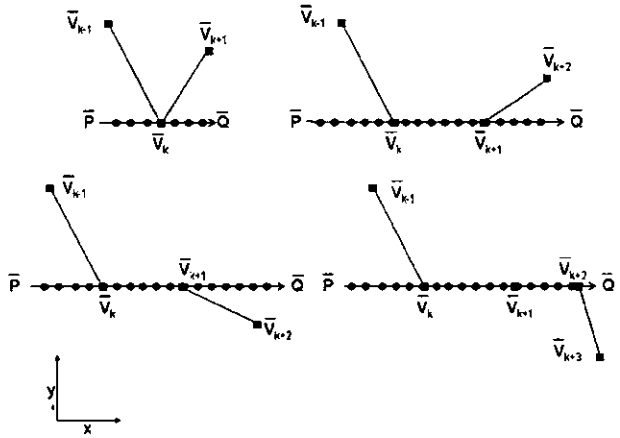


FIG. 3. Occurrences of the polygon vertex lying on the tracing ray $\overline{PQ}$. (top left) At $\bar{V}_k$ is assigned the opposite sign to $\bar{V}_{k-1}$. Consequently, a sign change is observed between $\bar{V}_{k-1}, \bar{V}_k$ and $\bar{V}_k, \bar{V}_{k+1}$ i.e., the ray $\overline{PQ}$ is observed to enter and exit the polygon at $\bar{V}_k$. (top right) The sign change is observed in this case between $\bar{V}_{k-1}, \bar{V}_k$ and also $\bar{V}_{k+1}, \bar{V}_{k+2}$ points from $\bar{V}_k$ to $\bar{V}_{k+1}$are located inside the polygon. Additionally, (bottom left) a sign change is observed at $\bar{V}_k$ where $\bar{V}_{k+1}, \bar{V}_{k+2}$, and $\bar{V}_{k+3}$ have the same sign as $\bar{V}_k$ and $\bar{V}_{k-1}$ has the opposite sign. Hence all points beyond $\bar{V}_k$ are inside the polygon. The similar situation, with an additional point on the ray $\overline{PQ}$ (bottom right) also registers all points beyond $\bar{V}_k$ as inside the polygon.

nique relies on a simple coordinate transformation to eliminate the cross-product computation in establishing intersections. Figure 2 shows an example of the polygon translation technique. The "trick" is to translate the polygon vertices and dose grid to a coordinate frame where $\bar{P}$ is at the origin and the ray $\overline{PQ}$ lies along a major axis, say the $X-$axis with Q situated on the opposite surface of the bounding box. Any point that extends beyond $\overline{PQ}$ is outside of the bounding box and is, therefore, not considered and a vertex's regional classification is then simply determined from the sign of its orthogonal coordinate, in this example the Y coordinate, in place of the sign of the cross product. For large numbers of vertices, making coordinate transformations is much less computationally expensive than performing many cross products. The remainder of the algorithm is the same as Siddon's algorithm where the cross-product signs are replaced with the coordinate signs. A similar method is described by Haines[9] and considers the individual points as the origin and traces a ray in the $+X$ direction counting the ray polygon intersections to effectively classify single points. A point known to be outside the polygon is selected as the origin using the bounding box placed on the data. All intersections along the $+X$ axis are obtained and all points along the ray classified at once. The implementation of such methods are fragile due to the need to handle the special cases (1) points lying on the polygon, (2) rays hitting a vertex, and (3) rays that are collinear with polygon edges. Suggested methods for addressing these issues[8] include an infinitesimal off set to the ray, polygon edge constraints and polygon reflections about the origin. However, these methods are inconsistent in their point classifications for some of the three special cases and additional checks have to be made. The method presented here avoids case (1) by requiring the starting point to be external to the polygon and deals with the other cases as follows.

Occurrences of the polygon vertex lying on the tracing ray $\overline{PQ}$ for both the Siddon and the polygon translation technique are demonstrated in Fig. 3. Note that a vertex, say $\bar{V}_k$, with a zero Y coordinate lies on the ray $\overline{PQ}$. If the preceding vertex V_{k-1} and the succeeding vertex V_{k+1} have different signs, the ray intersects the polygon at the vertex $\bar{V}_k$. If the preceding vertex $\bar{V}_{k-1}$ and the succeeding vertex $\bar{V}_{k+1}$ have the same sign, the ray intersects the boundary of the polygon at the single vertex $\bar{V}_k$ and either lies inside or outside the polygon right before and right after this vertex (counts as 2 intersections entering and leaving the point). However, if the Y coordinate of either $\bar{V}_{k-1}$or $\bar{V}_{k+1}$ (or both) is equal to zero as well, the ray $\overline{PQ}$ travels along the boundary of the polygon and extra care must be taken. In particular, there is a need to keep track of whether the ray $\overline{PQ}$ was outside or inside the polygon just before reaching the first vertex with zero Y coordinate to determine what happens when a vertex with nonzero Y coordinate is encountered. More formally, such cases can be dealt with as follows. If a vertex, say $\bar{V}_k$, is encountered with a zero Y coordinate a sign is associated with this coordinate that depends on the sign of the Y coordinate of the preceding vertex, $\bar{V}_{k-1}$. If the Y coordinate of $\bar{V}_{k-1}$ is negative then a positive sign is associated with the zero Y coordinate of $\bar{V}_k$, if the Y coordinate of $\bar{V}_{k-1}$ is positive a negative sign is associated with the zero Y coordinate of $\bar{V}_k$, and if the Y coordinate of $\bar{V}_{k-1}$ is zero then associate the same sign with the zero Y coordinate of $\bar{V}_k$. In addition, vertices are always permuted in such a way that the Y coordinate of the first vertex, $\bar{V}_1$, is nonzero. Then a change in sign denotes an intersection at the last vertex with a zero

value Y coordinate. Note that Siddon's algorithm only considers pairs of vertices and therefore is not able to deal correctly with such cases, which are often encountered in clinical radiotherapy data. For example, applying the algorithm to the instance found in the upper left of Fig. 3:

$$\text{if } \overline{V_k} = 0,$$
$$\text{if } \overline{V_{k-1}} = + \text{ then set } \overline{V_k} = -$$
$$\text{else if } \overline{V_{k-1}} = - \text{ then set } \overline{V_k} = +$$
$$\text{then if } \overline{V_k} \neq \overline{V_{k+1}} \text{ only the point at } \overline{V_k} \text{ is in the polygon}$$
$$\text{if } \overline{V_k} = \overline{V_{k+1}} \text{ all points beyond } \overline{V_k} \text{ are inside the polygon.}$$

B. Incremental ray tracing and virtual stereographic projection truncation for determination of radiological depth

The computer graphics literature contains numerous algorithms[12–23] for illumination of visible surface determination and volume rendering. Of these, ray surface intesection algorithms are the most applicable to radiological path determination and although some of these algorithms may have been implemented in treatment planning systems, they remain largely absent from the medical physics peer reviewed literature. Geometrically, voxel data can be considered as being defined by three orthogonal sets of parallel voxel boundary planes in three dimensions.[12,17,24] The choice of voxel representation depends upon the distribution of useful data through the object space. Regularly spaced 3D grid representations,[15] used for this investigation due to information requirements, are computationally simple to visualize and implement. They can, however, suffer from unnecessary ray-voxel intersection calculations in regions of low population density. Adaptive grids which aim to accelerate the traversal process by subdividing, the object space with data density[13,17,22,24] are available. These methods, although faster for small sparse problems, require a data structure, such as an octree,[25] that describes the connectivity of the voxel grid to determine ray-voxel intersections. In 1985, Siddon proposed the use of a linear-time algorithm that takes advantage of regular spaced 3D lattice-type representation of the patient voxel data for ray tracing. The radiological path must therefore be established for individual beamlets from the source to each voxel intersected by the radiation as shown in Fig. 4. In Siddon's implementation, iterative ray-tracing is performed from the point of the radiation source, say $\vec{S}$, to the center of each of the patient voxels, say $\vec{R}$, that lie within the beam path. Any point $\vec{T}$ along the ray $\overline{SR}$ can then be obtained from the parametric equation of a line given in Eq. (3) such that

$$\vec{T} = \vec{S} + \alpha(\vec{R} - \vec{S}), \tag{3}$$

where again α is the intersection parameter, and represents the fractional length along the ray $\overline{SR}$ where it intersects a voxel boundary plane. The calculation of all of the α values for a set of voxel data is performed independently for each orthogonal set of planes and is given by

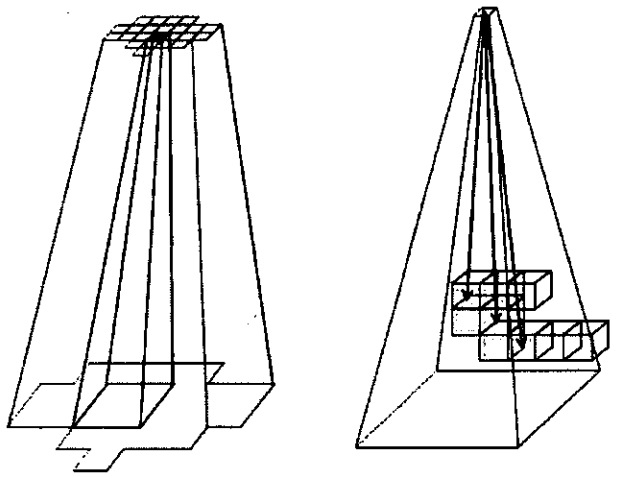


FIG. 4. Representation of the conformed beam as beamlets. The beam is considered as 1 cm^2 beamlets that can be blocked ("switched off") by multileaf-collimator leaves to conform the beam shape to that of the target (left). The radiological path is established for each beamlet-voxel-intersection (right) by ray-tracing from the source to the center of the voxels.

$$\alpha_X = \frac{X_{\text{plane}} - X_S}{X_R - X_S}, \tag{4}$$

$$\alpha_Y = \frac{Y_{\text{plane}} - Y_S}{Y_R - Y_S}, \tag{5}$$

and

$$\alpha_Z = \frac{Z_{\text{plane}} - Z_S}{Z_R - Z_S}, \tag{6}$$

where X_{plane}, Y_{plane}, and Z_{plane} are the orthogonal offsets of X-, Y-, and Z-planes and α_X, α_Y, and α_Z are the ray-plane intersection parameters, respectively. All of the orthogonal sets of intersection parameters must be concatenated into single set, sorted, and reduced to eliminate duplicate values. The indices of the intersected voxels are then obtained from the midpoints of neighboring α's and the path lengths are computed as the difference of points $\vec{T}_m$ and $\vec{T}_{m+1}$ where $m = 1, \ldots, N_\alpha - 1$ and N_α is the number of intersection parameters found. Intersected voxel indices are then used to extract the voxel density from the volumetric image data for calculation of the radiological path length. However, as reported by Jacobs *et al.*[26] a significant amount of computational time is required for the repeated calculation of voxel indices and the concatenation of α's into a single vector. Figure 5 shows a two-dimensional example of the representation of patient voxel space for radiological path determination.

In this work, an existing incremental voxel ray-tracing algorithm[26] is combined with a method of virtual stereographic projection to significantly reduce the computational cost of obtaining radiological path lengths. Several techniques are presented in the graphics literature which, deal with voxel traversal of a regular 3D grid. The differences between the various algorithms lies in the decision making process to establish which voxel the ray is currently moving through. Levoy's[13] technique uses evenly distributed steps

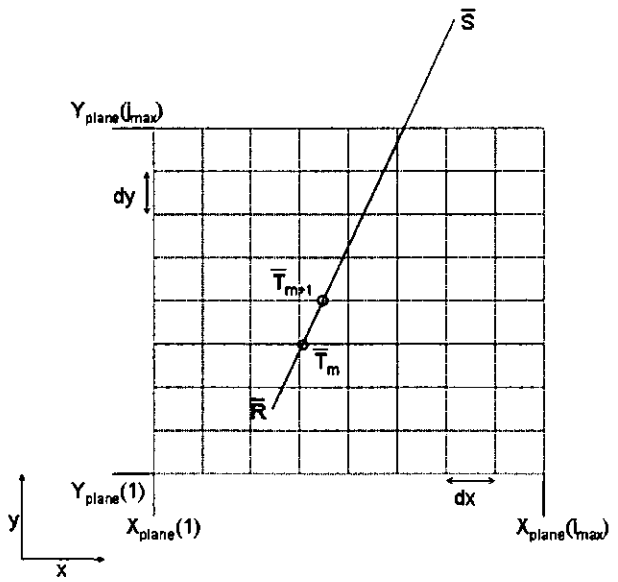


FIG. 5. Two-dimensional representation of the patient voxel spacing for radiological path determination. The ray is traced between $\bar{S}$ and $\bar{R}$ and the fractional length at which ray-plane intersections, $\overline{T_m}$ and $\overline{T_{m+1}}$, occur determined. The radiological path length is then obtained from the product of $|\overline{RS}|$ and the difference of consecutive α's and the voxel density.

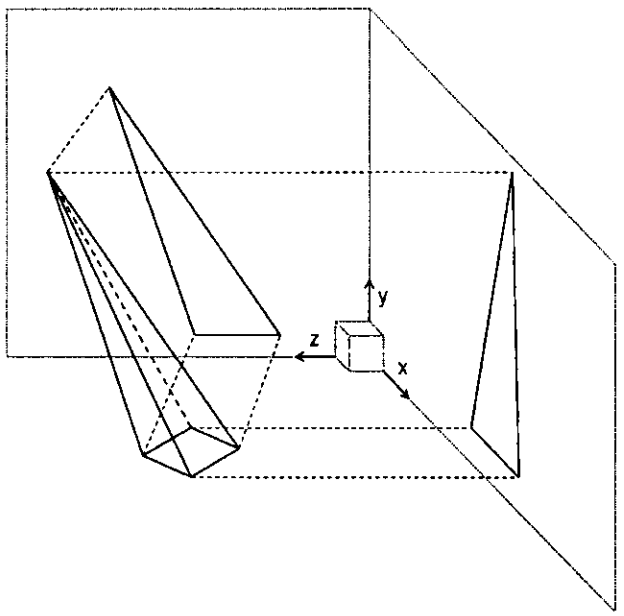


FIG. 6. Stereographic projection of a tetrahedral beamlet. The convex hull of the beamlet is projected onto the XY and YZ planes along with the dose grid. Any grid point contained within the beamlet projection in both the XY and YZ planes is contained intersected by the 3D beamlet.

along the ray. The ray tracing implemented by Siddon[2] obtains intersections along each of the three axes before combining the results while incremental methods generally deal only with one axis at a time.[15] Methods of incremental voxel ray tracing have been employed in Monte Carlo simulations[27–30] where it is required for electron transport and often used with the δ-scattering method of Woodcock[31] for photon transport to avoid calculating the full radiological path length of a photon. Incremental voxel ray-tracing methods have also been employed as a natural fast ray tracing technique for nuclear medicine image reconstruction applications, where nuclear decays occur within the patient volume and radiations travel out to the detector system.[26] Fujimoto *et al.*,[15] developed the 3D digital differential analyzer (3DDDA) for incremental ray tracing. The method depends on a driving axis (DA) and an error term calculated perpendicular to the DA for the ray-voxel intersection process. The present method eliminates the need for either of these in performing the incremental process. To keep track of voxel indices, the voxel of interest $\bar{R}$ is taken as the starting point and ray tracing is performed back to the source $\bar{S}$. Thus, the initial voxel indices are known and, due to the incremental nature of the algorithm, the indices of all inter-

TABLE I. Comparison of the Siddon algorithms with the methods from this work for five head and neck and five prostate cases.

Case	Number of voxels ($\times 10^6$)	Number of beamlets	Number of targets	Number of critical structures	Fold improvement for voxel classification	Fold improvement for voxel truncation	Radiological path (fold improvement)
				Head and Neck			
1	1.71	1466	2	12	100	1.8	1.8
2	1.61	1493	2	8	132	1.8	2.8
3	1.34	1406	2	8	148	1.8	2.9
4	1.30	1184	2	8	114	1.9	3.0
5	1.99	2398	4	14	192	1.8	2.9
				Prostate			
6	1.06	486	1	9	339	2.0	2.7
7	1.16	457	1	6	360	2.0	2.6
8	1.03	543	1	7	257	1.9	2.6
9	0.58	352	1	5	174	1.7	2.8
10	0.78	234	1	5	194	2.0	2.9

TABLE II. Comparison with Siddons techniques for the average calculation times per beamlet for the virtual stereographic projection truncation and incremental ray tracing algorithms. The average number of voxels per beamlet was ~15 600 and ~33 600 for head and neck and prostate cases, respectively.

	Siddons Algorithms		Present Algorithms	
Case	Voxel truncation (sec per beamlet)	Radiological path (sec per beamlet)	Voxel truncation (sec per beamlet)	Radiological path (sec per beamlet)
		Head and Neck		
1	0.23	3.46	0.13	1.97
2	0.23	4.03	0.13	1.42
3	0.21	3.80	0.11	1.32
4	0.20	3.33	0.10	1.13
5	0.27	4.53	0.15	1.55
		Prostate		
6	0.19	7.85	0.10	2.91
7	0.21	8.36	0.10	3.21
8	0.18	7.41	0.10	2.85
9	0.11	5.63	0.06	2.00
10	0.18	7.60	0.09	2.60

sected voxels are acquired simultaneously with the determination of the intersection path lengths, hence avoiding the inefficiencies of the Siddon algorithm due to repeated calculation of voxel indices and the concatenation of α's. Figure 5 demonstrates a 2D example of incremental voxel ray tracing. For the purposes of dose computation, $\vec{R}$ lies at the center of a voxel (i,j) and the radiological path is required along $\overline{RS}$ for each voxel. The first boundary intersection along the ray $\overline{RS}$, is obtained by calculating the three α's for the nearest X,Y, and Z voxel bounding planes in the direction of the ray and taking the minimum of these as the next intersection. The voxel index is incremented in the direction of the plane corresponding to the minimum of the three α's. Then the next α in the same direction is computed and the minimum of this and the remaining two α's is used to determine the next actual intersection. This process is repeated toward $\vec{S}$ until $\vec{S}$ is reached, the ray leaves the voxel data, or the ray leaves the patient volume.

TABLE III. Performance times per beamlet for voxel truncation using the beamlet extent approach and the virtual stereographic projection method.

Case	Stereographic projections (sec per beamlet)	Beamlet extent (sec per beamlet)	Improvement (fold)
		Head and Neck	
1	0.13	2.17	17.18
2	0.13	2.03	15.58
3	0.11	1.65	14.50
4	0.10	1.52	14.99
5	0.15	1.85	19.61
		Prostate	
6	0.10	1.07	11.19
7	0.10	2.51	24.00
8	0.10	2.43	25.57
9	0.06	1.27	20.94
10	0.09	1.85	19.61

To further improve the efficiency of the approach the voxels of interest $\vec{R}$ that would result in extremely small dose values were truncated using two separate methods. In the beamlet extent approach a coordinate transformation of the entire voxel grid is performed for each beamlet so that the direction of the beamlet is aligned with a major coordinate axis, i.e., the Y axis. Voxels outside of a projected radius or box that encloses the beamlet are then truncated to avoid computing extremely small dose values. Further truncation is achieved by ray tracing to the extent of the beamlet and obtaining the points of intersection with the CT slices that contain 2D structure polygons. Point classification then determines voxels to be above (outside the beamlet) or below or on (inside the beamlet) the tracing ray and the coordinates of the CT intersections bound the voxels in X and Z. With this method the coordinate transformation represents a large proportion of the computation time required. To avoid this costly coordinate transformation, each beamlet was modeled as a 3D tetrahedron that enclosed the geometric projection of each beamlet with a lateral margin sufficient to capture any scatter dose greater than a part-per-hundred-million from the primary photon beam in our finite sized pencil beam dose model. For the 6 MV 1×1 cm^2 pencil beams used in this study the margin was 1.5 cm (creating a 4×4 cm^2 area at iso-center). To rapidly determine which voxels where encased inside of the beamlet tetrahedron margin, projected the vertices of the 3D tetrahedron onto two orthogonal voxel boundary planes to generate the convex hull[32] of 2D polygons in each plane. An example of this projection is shown in Fig. 6. Then the coordinate translation polygon ray-tracing algorithm was used to classify those voxels inside of both projected 2D polygons.

III. RESULTS

To compare the present ray-tracing approach with the existing techniques, both the algorithms of Siddon and the algorithms presented above were implemented in the University of Florida optimized radiation therapy (UFORT) treatment planning decision support system (TPDSS). The UFORT TPDSS was developed using a commercial technical programming language (Matlab, Mathworks, Inc.) and was run on a 3.6 GHz threading enabled Pentium IV processor with 2 GB of RAM. This system was designed to accept DICOM communications of treatment planning image, structure, and plan data from commercial treatment planning or simulation software. Ten cases of 3D image-based treatment-planning data from two different sites (five head-and-neck and five prostate cases) were used for testing the algorithms. The system anonymized the patient data for research purposes and converted the data to an internal data format. To compute dose, a simple empirical finite sized beamlet model was employed that has been described elsewhere.[33] These planning cases required the classification of up to 2 million voxels on a 2.5 mm isotropic dose grid into targets and structures represented as extruded polygons and the computation of voxel dose was required for between 350 and 2400 beamlets per case. For head and neck cases an average of ~15 600 voxels-per-beamlet were observed while for prostate cases the number was ~33 600. Note that in our finite sized beamlet model ray tracing is performed to every voxel in each beamlet. The case sizes and resulting improvements, over the Siddon algorithms, are given in Table I. As described above, the polygon ray-tracing technique has two implementations as part of the treatment planning process: firstly in voxel classification i.e., assigning voxels to structures, and secondly for voxel truncation (determination of those voxels intersected by a beamlet). The translation polygon ray-tracing technique shows a 100–360 (213–2.13 s for a typical head and neck case and 590 s down to 1.64 s for a typical prostate case) fold improvement in computational speed over the Siddon method for the voxel classification of biological structures in head and neck and prostate cases. Voxel truncation using virtual stereographic projection with the translation technique was observed to reduce the computation period by 1.7–2.0 fold. This is equivalent to a reduction from 0.2 to 0.11 s per beamlet, averaged from Table II, when compared to the method of Ref. 1 for a 7 beam plan with equidistant spacing and a dose grid of resolution 2.5 mm. Table III demonstrates the improvement of the stereographic projection method, per beamlet, compared to the beamlet extent approach for voxel truncation. Calculation times are reduced by up to 25 fold per beamlet, which for the current investigation translates to total time reductions, for all beamlets from ~1800–4400 to 118–360 s for head and neck cases and from ~432–1320 to 21–55 s for prostate. The incremental ray-tracing method showed a 2.6–3.1 fold improvement in computational speed compared to Siddon's algorithm.[2] Calculation times, averaged from Table II, were observed to fall from 3.8 to 1.3 s per beamlet for head and neck cases and from 7.6 to 2.7 s in the prostate cases.

IV. CONCLUSIONS

The works of Siddon have been widely employed in image reconstruction and dose computation since their introduction in the mid 80's and continue to find widespread use with over 130 citations for these works at the present time. The algorithms presented, as a whole, show significant improvement over the conventional techniques introduced by Siddon. However, if the intra-fractional reassessment of treatment plans for adaptive IGIMRT is to be realized considerable speed up of the current methods will still be required. As each 1×1 cm^2 beamlet can be considered separately the approach is trivially parallelized. Consequently, there is a natural inclination to migrate to parallel computation to significantly further reduce the time necessary for plan calculation for adaptive IMRT FMO.

ACKNOWLEDGMENTS

This work was supported in part by NSF Grant No. DMI-0457394, NCI Grant No. RO1 CA 100636, and the State of Florida DOH Grant No. 04NIR-03-4664.

[a]Author to whom correspondence should be addressed. Telephone: 352-265-8217; Fax: 352-265-8417. Electronic mail: dempsey@ufl.edu

[1]R. L. Siddon, "Prism representation: A 3D ray-tracing algorithm for radiotherapy applications," Phys. Med. Biol. **30**(8), 817–824 (1985).
[2]R. L. Siddon, "Fast calculation of the exact radiological path for a three-dimensional CT array," Med. Phys. **12**(2), 252–258 (1985).
[3]D. Yan, F. Vicini, J. Wong, and A. Martinez, "Adaptive radiation therapy," Phys. Med. Biol. **42**(1), 123–132 (1997).
[4]D. Yan, F. Vicini, J. Wong, and A. Martinez, "Adaptive modification of treatment planning to minimize the deleterious effects of treatment setup errors," Int. J. Radiat. Oncol., Biol., Phys. **38**(1), 197–206 (1997).
[5]C. Wu, R. Jeraj, W. Lu, and T. R. Mackie, "Fast treatment plan modification with an over-relaxed Cimmino algorithm," Med. Phys. **31**(2), 191–200 (2004).
[6]C. Wu, R. Jeraj, W. Lu, and T. R. Mackie, "Re-optimization in adaptive radiotherapy," Phys. Med. Biol. **47**(17), 3181–95 (2002).
[7]M. Birkner, D. Yan, M. Alber, J. Liang, and F. Nusslin, "Adapting inverse planning to patient and organ geometrical variation: Algorithm and implementation," Med. Phys. **30**(10), 2822–31 (2003).
[8]J. O'Rourke, *Computational geometry in C*, 2nd ed (Cambridge University Press, Cambridge, 1998).
[9]*Point in polygon strategies*, edited by P. S. Heckbert (Editor), Chapter by E. Haines (Graphics Gems IV, Morgan Kaufmann, San Francisco, CA, 1994), pp. 26–45.
[10]S. Green, *Parallel processing for computer graphics* (Cambridge University Press, Cambridge, 1991), pp. 6–22.
[11]A. Nassar, P. Walden, E. Haines, T. Dickens, R. Capelli, S. Narasimhan, C. Jam, and S. MacMartin, "Fastest Point in Polygon Test," Ray Tracing News, 5(3) (1992): Archive at http://www.raytracingnews.org/
[12]A. S. Glassner, *Ray Tracing* (Graphics Gem, Morgan Kaufmann, San Francisco, CA, 1990), pp. 385–396.
[13]M. Levoy, "Efficient ray tracing of volume data," ACM Trans. Graphics **9**(3), 245–261 (1990).
[14]N. Stolte and R. Caubet, "Descrete ray-tracing of huge voxel spaces," Comput. Graph. Forum **14**(3), 383–394 (1995).
[15]A. Fujimoto, T. Tanake, and K. Iwata, "ARTS: Accelerated ray-tracing systems," IEEE Comput. Graphics Appl. **6**(4), 16–26 (1986).
[16]M. R. Stytz, G. Frieder, and O. Frieder, "Three-dimensional medical imaging: Algorithms and computer systems," ACM Comput. Surv. **23**(4), 421–499 (1991).
[17]A. Kaufman, D. Cohen, and R. Yagel, "Volume Graphics," Computer **26**(7), 51–64 (1993).
[18]R. Yagel, D. Cohen, and A. Kaufman, "Discrete ray tracing," Plan Canada **12**(5), 19–28 (1992).
[19]J. S. Pantazopoulos and S. G. Tzafestas, "An efficient algorithm for ray

tracing," J. Intell. Robotic Syst. **28**, 171–180 (2000).
[20]S. Parker, M. Parker, Y. Livant, P. P. Sloan, and C. Hansen, "Interactive ray tracing for volume visualization," IEEE Trans. Vis. Comput. Graph. **5**(3), 238–250 (1999).
[21]T. L. Kay and J. T. Kajiya, "Ray tracing complex scenes," Comput. Graphics **20**(4), 269–278 (1986).
[22]K. S. Klimaszewski and T. W. Sederberg, "Faster ray tracing using adaptive grids," IEEE Comput. Graphics Appl. **17**(1), 42–51 (1997).
[23]A. S. Glassner, "Space subdivision for fast ray tracing," IEEE Comput. Graphics Appl. **4**, 15–22 (1984).
[24]J. M. Snyder and A. H. Barr, "Ray tracing complex models containing surface tessellations," Comput. Graphics **21**(4), 119–128 (1987).
[25]I. E. Sutherland, R. F. Sproull, and R. A. Schumacker, "A characterization of ten hidden surface algorithms," ACM J. Computing Surveys **6**(1), 1–55 (1974).
[26]F. Jacobs, E. Sundermann, B. De Sutler, M. Christiaens, and I. Lemahieu, "A Fast Algorithm to Calculate the Exact Radiological Path Through Pixel or Voxel Space," Journal of Computing and Information Technology, CIT **6**(1), 89–94 (1998).
[27]D. W. O. Rogers and A. F. Bielajew, "Monte Carlo techniques of electron and photon transport for radiation dosimetry" in editors, *The Dosimetry of Ionizing Radiation*, edited by K. R. Kase, B. E. Bjärngard, and F. H. Attix, (Academic, New York, 1990), Vol. III, pp. 427–539.
[28]J. J. DeMarco, T. D. Solberg, and J. B. Smathers, "A CT-based Monte Carlo simulation tool for dosimetry planning and analysis," Med. Phys. **25**(1), 1–11 (1998).
[29]J. Sempau, S. J. Wilderman, and A. F. Bielajew, "Fast, accurate Monte Carlo code optimized for photon and electron radiotherapy treatment planning dose calculations," Phys. Med. Biol. **45**(8), 2263–91 (2000).
[30]C. N. Zeeb, J. S. Dolaghan, and P. J. Burns, "An efficient monte carlo particle tracing algorithm for large, arbitary geometries," Numer. Heat Transfer, Part B **39**, 325–344 (2001).
[31]E. Woodcock, T. Murphy, P. Hemmings, and S. Longworth, "Techniques used in the GEM code for Monte Carlo neutronics calculations in reactors and other systems of complex geometry," In *Proceedings of the Conference on Applications of Computing Methods to Reactor Problems*, p. 557 (1965).
[32]C. B. Barber, D. P. Dobkin, and H. T. Huhdanpaa, "The Quickhull Algorithm for Convex Hulls," ACM Trans. Math. Softw. **22**(4), 469–483 (1996).
[33]H. E. Romeijn, R. K. Ahuja, J. F. Dempsey, A. Kumar, and J. G. Li, "A novel linear programming approach to fluence map optimization for intensity modulated radiation therapy treatment planning," Phys. Med. Biol. **48**(7), 3521–3542 (2003).

 3382

Phys. Med. Biol. 45 (2000) 2263–2291. Printed in the UK PII: S0031-9155(00)12211-0

DPM, a fast, accurate Monte Carlo code optimized for photon and electron radiotherapy treatment planning dose calculations

Josep Sempau†‡, Scott J Wilderman† and Alex F Bielajew†

† Department of Nuclear Engineering and Radiological Sciences, The University of Michigan, Ann Arbor, MI, USA

‡ Institut de Tècniques Energètiques, Universitat Politècnica de Catalunya, Diagonal 647, 08028 Barcelona, Spain

Received 29 February 2000

Abstract. A new Monte Carlo (MC) algorithm, the 'dose planning method' (DPM), and its associated computer program for simulating the transport of electrons and photons in radiotherapy class problems employing primary electron beams, is presented. DPM is intended to be a high-accuracy MC alternative to the current generation of treatment planning codes which rely on analytical algorithms based on an approximate solution of the photon/electron Boltzmann transport equation. For primary electron beams, DPM is capable of computing 3D dose distributions (in 1 mm^3 voxels) which agree to within 1% in dose maximum with widely used and exhaustively benchmarked general-purpose public-domain MC codes in only a fraction of the CPU time. A representative problem, the simulation of 1 million 10 MeV electrons impinging upon a water phantom of 128^3 voxels of 1 mm on a side, can be performed by DPM in roughly 3 min on a modern desktop workstation. DPM achieves this performance by employing transport mechanics and electron multiple scattering distribution functions which have been derived to permit long transport steps (of the order of 5 mm) which can cross heterogeneity boundaries. The underlying algorithm is a 'mixed' class simulation scheme, with differential cross sections for hard inelastic collisions and bremsstrahlung events described in an approximate manner to simplify their sampling. The continuous energy loss approximation is employed for energy losses below some predefined thresholds, and photon transport (including Compton, photoelectric absorption and pair production) is simulated in an analogue manner. The δ-scattering method (Woodcock tracking) is adopted to minimize the computational costs of transporting photons across voxels.

1. Introduction

Several researchers have recently suggested that Monte Carlo (MC) based systems will soon become the dominant vehicles for dose computation in radiotherapy treatment planning (Bielajew 1994a, 1997, Mohan 1997, Hartmann-Siantar *et al* 1997). The superior accuracy of the MC method, which converges to results which are exact to the degree to which physical parameters are known, over that of deterministic models has long been well established. Public domain codes such as EGS4 (Nelson *et al* 1985, Bielajew *et al* 1994), ITS (Halbeib 1989, Halbeib *et al* 1992), MCNP (Briesmeister 1993), and PENELOPE (Baró *et al* 1995, Salvat *et al* 1996, Sempau *et al* 1997) have all been extensively benchmarked against experimental data for a wide range of materials and energies. EGS4 in particular has been thoroughly tested in the specific region of dosimetric interest (Rogers and Bielajew 1989b, 1990), and is widely accepted as a computational standard for radiotherapy dose

0031-9155/00/082263+29$30.00

calculations. Further, the near equivalence of the prevalent MC codes has also been established (Rogers and Bielajew 1989a, Andreo 1991), and their differences shown to be of little significance in the radiotherapy dose calculation problem†.

By contrast, the deterministic algorithms currently used for calculating electron dose in treatment planning systems rely on analytic approximations to the solution of the transport equation which fail to adhere to their limiting conditions in certain radiotherapy applications. Errors up to 50% for electron beams (Cygler *et al* 1987) and up to 30% for problems involving photon and electron transport near inhomogeneities (Ma *et al* 1999) have been reported. These discrepancies arise because the deterministic methods are based on approximate analytical solutions of the transport problem in semi-infinite media which are then modified semiempirically to account for inhomogeneities. Such methods are often not adequate for treatment planning computations in which interfaces between materials with large differences in density and/or atomic numbers (e.g. soft tissue, bone and air) play an important role. MC based techniques, on the other hand, are capable of modelling heterogeneities with a fine granularity.

Thus far, the impediment to the widespread implementation of MC based methods for dose computation has been that, even with continuing advances in computer architecture and clock speed, the currently available codes are quite slow. The practical requirement imposed by clinical radiotherapy treatment planning systems is to provide dose distributions of sufficient accuracy (~2–3% of the dose maximum) within a time of practical clinical relevance ($\lesssim$5 min) and with a modest investment in computer hardware. Though most MC programs are sufficiently fast for simulating dose deposition in homogeneous media and simple geometries, radiotherapy applications involve numerous variations of material *and* density over small distances. Patient geometry is usually simulated as a map of densities over a large number (128^3) of relatively small (~1–4 mm) parallelepipeds (voxels), obtained from computed tomography (CT) scans. Currently, Monte Carlo simulation of absorbed dose for such large-scale problems is feasible only when employing computer resources of a scale not generally available in medical centres (Hartmann-Siantar *et al* 1995, 1997, Ma *et al* 1999).

Recently, Keall and Hoban (1996), Neuenschwander and Born (1992), Neuenschwander *et al* (1995), Kawrakow *et al* (1996) and the PEREGRINE code (Hartmann-Siantar *et al* 1995, 1997) have attempted to surmount the CPU constraint by significantly modifying the basic MC electron transport algorithm. These new methods rely on some combination of simplifying the physics to different degrees of accuracy; reusing all or parts of particle histories; and/or implementing parallel processing (requiring a significant investment in hardware). In this work we present a new MC algorithm, the 'dose planning method' (DPM) for the simulation of coupled electron–photon transport in radiotherapy treatment planning without reliance on these limiting approximations and requirements.

DPM employs the standard condensed history model for electron transport, and falls into what has been called by Berger (1963) a 'mixed' scheme for the treatment of energy losses, treating large energy transfer collisions in an analogue sense and using the continuous slowing down approximation (CSDA) to model small-loss collisions. Gains in performance derive from a series of significant enhancements to the algorithm for transporting particles from point to point (the 'transport mechanics') and corresponding reformulation of the distribution functions describing the physics, as described below.

The first modification involves the employment and refinement of a new step size independent multiple-scattering theory (Kawrakow and Bielajew 1998b). Our method

† The comparisons performed by Rogers and Bielajew employed the ETRAN code (Berger 1963, Seltzer 1989, 1991) from which the electron transport physics (with some subtle modifications) for ITS and MCNP was derived.

is a robust implementation of the Lewis (1950) formulation of Goudsmit–Saunderson theory (Goudsmit and Saunderson 1940a, b). Both provide exact solutions for the angular distribution of an electron traversing a given distance. However, the Lewis formulation allows the cross section to vary along the path according to the CSDA energy-loss model. The underlying cross section used here is the screened Rutherford cross section with the Molière screening factor (Molière 1947), after that implemented in Molière's small-angle multiple-scattering theory (Molière 1948) and employed in the EGS4 code with a correction for large angle suggested by Bethe (1953). From the validity of the EGS4 code for radiotherapy applications, we infer that the use of this form is sufficiently accurate. In addition to being derived under an exact framework, because it can be recast into a form independent of energy, the Kawrakow–Bielajew multiple scattering formulation provides a vehicle which can be exploited to permit transport across inhomogeneities, as discussed later. Use of the Kawrakow–Bielajew distribution provides one other advantage. Larsen (1992) has demonstrated that the accuracy of a transport simulation scheme for charged particles is dominated by the faithfulness of the multiple-scattering theory it employs. In the limit of small electron step size, the correct solution to the transport equation is guaranteed *if and only if* the multiple-scattering theory faithfully reproduces the discrete single-scattering distributions. To our knowledge, the Kawrakow–Bielajew formalism yields the only purely multiple-scattering distribution function which is correct in the multiple-, plural- and single-scattering regimes. Thus, DPM is guaranteed to always converge to a correct solution as the step size is reduced.

The second major innovation introduced by DPM lies in the use of new transport mechanics, i.e. the algorithm for moving charged particles from point-to-point in media given the composition of the material traversed, the length of the step and the multiple-scattering angle. Larsen's work suggests that the schemes employed in current general-purpose codes are not optimal, and that a measure of the quality of a transport mechanism is measured by how quickly it converges to the small step size limit. Transport schemes can be characterized also by their adherence to the exact spatial-angular moments first reported by Lewis (1950), and a new scheme with high-order convergence has been reported recently (Kawrakow and Bielajew 1998a). The implementation of this new method is computationally and algorithmically demanding, however, and so DPM has adopted the 'random hinge' scheme† employed in PENELOPE. This algorithm has been shown to be almost as accurate as the Kawrakow–Bielajew (1998a) transport mechanics in preserving the basic Lewis moments, but has a much simpler implementation. Recent work of Larsen‡ and of Bielajew and Salvat (2000) demonstrates that higher-order convergent schemes exist, but they were not studied due to algorithmic complexity.

A third new technique introduced by DPM is the use of large electron transport steps, in which many voxels may be traversed before sampling a multiple scattering angle. This is made possible because of the stability of the random hinge algorithm across heterogeneities, the accuracy of the Kawrakow–Bielajew distribution, and because the multiple-scattering angle, when the step size is suitably scaled in terms of energy and scattering in the medium (as described below), is very nearly independent of atomic number. This feature of multiple scattering distributions has been shown to be rigorously true in small-angle theory (Bothe 1921a, b, Wentzel 1922, Molière 1948, Bielajew 1994b). We thus assume that the small residual dependences on the media of both the scattering distributions and of the random hinge Lewis moments can be safely ignored for radiotherapy-class problems, and relatively large steps (of the order of 5 mm) can be employed.

† We are grateful to Dr Ronald Kensek of Sandia National Laboratories for this colourful nomenclature.

‡ We are grateful to Dr Ed Larsen of the University of Michigan for providing us with this information ahead of publication.

Because the differential cross section is a fairly strong function of energy and there is significant energy loss over the long steps taken in DPM, a fourth modification has been introduced which scales the step sizes by the number of (material and energy dependent) transport mean free paths traversed. This preserves the total amount of scattering modelled by the multiple-scattering distribution functions over the long steps, and is essential for permitting tracking across sharp heterogeneities.

In these as well as other features, DPM exploits the small dynamic range (in energy and material) of radiotherapy class problems. Energies are limited to those between ~100 keV and ~20 MeV, and, while the program has been benchmarked against a wide range of atomic numbers for completeness, because in most clinical applications only a few low-atomic-number materials are seen, certain cross sections and distribution functions are determined by scaling them appropriately to exactly computed data for water.

In the following sections, we present in detail the multiple-scattering model, electron transport mechanics, treatment of large energy loss processes, the photon transport algorithm, and cross-voxel transport found in DPM. Results from electron dose deposition simulations in homogeneous and inhomogeneous phantoms, as well as in a CT geometry, are then presented, followed by a section devoted to the analysis of CPU run time and of the simulation efficiency achieved.

2. Multiple scattering

In all condensed history MC programs, the effect of the large number of elastic interactions which occur over a given pathlength is modelled by means of a multiple scattering theory. The theory of Goudsmit and Saunderson (1940a) (GS hereafter), which is exact if the cross section is constant over the step, describes the angular deflection of electrons after travelling a given pathlength s in terms of transport coefficients g_ℓ ($\ell = 0, 1 \ldots \infty$), defined by

$$g_\ell = 1 - \int_{-1}^{1} \mathrm{d}\omega \, P_\ell(\omega) p(\omega). \tag{1}$$

Here $\omega = \cos\theta$ is the angular deviation with respect to the initial electron direction, P_ℓ is the ℓth Legendre polynomial and the quantity

$$p(\omega) \equiv \frac{\sigma(\omega)}{\int_{-1}^{1} \mathrm{d}\omega' \, \sigma(\omega')} \tag{2}$$

represents the probability density function (PDF) associated with the single-event differential cross section (DCS) $\sigma(\omega)$. Under this formalism, the angular distribution of electrons having traversed a distance s is given by

$$F_{\mathrm{GS}}(\omega)\,\mathrm{d}\omega = \sum_{\ell=0}^{\infty} \left(\ell + \frac{1}{2}\right) P_\ell(\omega) \exp\left(-\frac{s}{\lambda} g_\ell\right) \mathrm{d}\omega \tag{3}$$

where λ is the elastic scattering mean free path (MFP) $1/N\sigma$, given that N is the atom density of the medium and σ is the total microscopic cross section. An important consequence of this formalism is that the transport coefficients fully characterize the elastic scattering process. Note that this series diverges for $\omega = 1$, because of the presence of uncollided particles. Numerically, this divergence appears as an instability in F_{GS} when values of s/λ (the number of MFPs along s) fall below 100.

As the GS model is exact only in as much as the transport coefficients are exact, $\sigma(\omega)$ must be chosen carefully. Scattering from a screened Rutherford potential, though not rigorously

accurate, provides a physically sound model for single elastic collisions, and leads to a PDF

$$p_R(\omega) = \frac{2\eta(1+\eta)}{(1+2\eta-\omega)^2} \tag{4}$$

where η is related to the screening parameter in the Rutherford potential. An advantage of using the screened potential PDF is that analytical expressions for all of the transport coefficients can be derived in terms of η. Further, η can be arbitrarily set so as to reproduce the first transport coefficient g_1 obtained from numerical integration of more accurate (and computationally cumbersome) DCSs found elsewhere (Mayol and Salvat 1997). Typical values of η for water in the energy range relevant for radiotherapy fall in the interval from 10^{-8} to 10^{-4}. In practice, a thousand coefficients are enough to ensure the convergence of the GS series except for very small pathlengths, and the approximate small momentum method of Kawrakow and Bielajew (1998b) can be used to calculate these coefficients.

Direct use of the screened Rutherford cross section and the GS theory, however, requires an impractical amount of computer memory to store accurate numerical representations of the resulting steeply forward-peaked distribution. Following Kawrakow and Bielajew (Kawrakow and Bielajew 1998b), this difficulty can be overcome by a change to a new angular variable u, which is defined so that the relation

$$\left|\frac{du}{d\omega}\right| = \frac{2B(1+B)}{(1+2B-\omega)^2} \tag{5}$$

is fulfilled, where B represents a free parameter that is called the 'broad screening' parameter, to associate it with the screened Rutherford shape. Moreover, if the boundary conditions

$$u(\omega = 1) = 0 \qquad \text{and} \qquad u(\omega = -1) = 1 \tag{6}$$

are imposed, equation (5) together with (6) fully determine u, and the expression for the new variable u is found to be

$$u = (1-\omega)\frac{1+B}{1+2B-\omega}. \tag{7}$$

The PDF $q(u)$ of the new variable is related with the GS distribution through

$$F_{GS}(\omega) \equiv q(u)\left|\frac{du}{d\omega}\right| = q(u)\frac{2B(1+B)}{(1+2B-\omega)^2} \tag{8}$$

and so q can be interpreted as a 'correction' factor that transforms a screened Rutherford PDF into the GS distribution. Equation (8) also shows that the sampling of ω values according to the PDF $F_{GS}(\omega)$ can be readily carried out using the rejection method, employing $q(u)$ as the rejection function.

2.1. Optimizing q

Rejection sampling for ω will be accurate and efficient and will require a manageable amount of computer memory only if $q(u)$, which depends only on s/λ, η and B, is sufficiently smooth. The introduction of the arbitrary parameter B in equation (8) provides a degree of freedom that can be exploited to manipulate the shape of q to bring this about. The requirement that q be as smooth as possible can be expressed mathematically as

$$\frac{\partial}{\partial B}\int_0^1 du[1-q]^2 = 0. \tag{9}$$

Since $q(u)$ is a PDF, its integral is 1 and the former equation simplifies to

$$\frac{\partial}{\partial B}\int_0^1 du\, q^2 = 0 \tag{10}$$

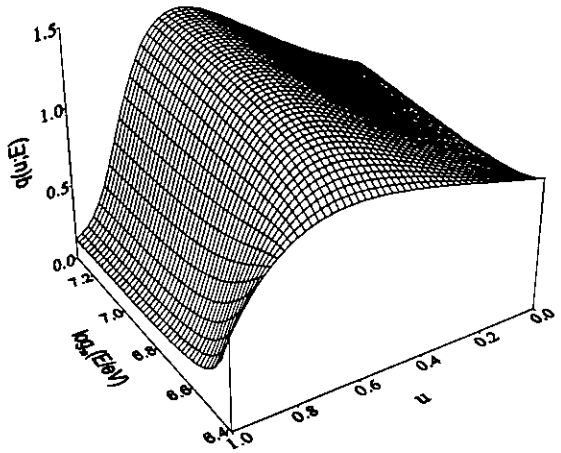


Figure 1. q surface for $s = 1$ cm as a function of the angular variable u and the logarithm of the kinetic energy in eV. Note that the vertical scale is linear. The lower limit E_L of the energy interval considered is chosen to be slightly larger than energy E_r for which the residual range is 1 cm. At energies below E_r, the surface collapses to the plane $q = 1$, and the parameter B in (8) goes to ∞, yielding an F_{GS} of 1/2, which implies isotropic scattering. To avoid this singularity in the interpolating routine, one must cut off at a slightly greater value, E_L. This, of course, implies the need for a last-step strategy, that is, the last step is treated as a special case.

which can be solved analytically (Kawrakow and Bielajew 1998b) to obtain B as a function of s/λ and η.

A fairly good approximation for $B(s/\lambda, \eta)$ (and for the corresponding q) can be obtained in the small-angle limit, i.e. when s/λ is not very large and the scattering is weak (equivalent to η being small). It is noted that within this approach, which has been extensively studied by Bielajew (1994b), neither q nor the ratio B/η depend on η, making the interpolation process somewhat simpler. However, as large step sizes are required to significantly reduce computation time in DPM, the conditions necessary to apply this simplification will not be met, and a numerical solution of the exact expression in (10) is used instead.

Since s, as discussed in a later section, is almost always expressed as a function of the kinetic energy of the electron, E, and since λ and the screening parameter η are also functions of E, q depends *only* on the dynamic variables u and E. In figure 1 a plot of the surface $q(u; E)$ is presented for $s = 1$ cm, showing that the change of variable introduced in equation (7) does indeed produce a smooth q. Typically, less than 20 kB (varying slightly with the selected $s(E)$) are needed to reproduce q with a mean accuracy better than 0.1%. The problem of sampling the GS distribution has thus been reduced to interpolating $q(u; E)$.

2.2. *Multiple scattering with energy losses*

Since electrons lose energy continuously as they pass through matter and the elastic scattering cross section is a fairly strong function of the electron energy, there is dependence on energy in both λ and g_ℓ in the exponential in (3). Lewis (1950) first accounted for this by recasting s as an integral over energy loss in the continuous slowing down approximation, noting that the pathlength can be expressed in terms of energy loss as

$$s = \int_{E-\Delta E}^{E} \frac{\mathrm{d}E}{S(E)} \equiv R(E) - R(E - \Delta E). \tag{11}$$

Here $S(E)$ is the energy-dependent energy loss per pathlength or CSDA stopping power, and $R(E)$ is called the CSDA range. The average energy loss ΔE for an electron with initial energy E travelling a given distance s can be determined by inverting the CSDA range, as in

$$\Delta E = E - R^{-1}(R(E) - s). \tag{12}$$

Thus Lewis was able to introduce energy dependence into (3) by using (11) to write $s/\lambda g_\ell$ in the exponential as an integral over energy loss

$$F_{\mathrm{L}}(\omega) = \sum_{\ell=0}^{\infty} \left(\ell + \frac{1}{2}\right) P_\ell(\omega) \exp\left(-\int_{E-\Delta E}^{E} \mathrm{d}E \frac{G_\ell}{S}\right) \tag{13}$$

in which

$$G_\ell \equiv \frac{g_\ell}{\lambda} \tag{14}$$

is defined as the ℓth inverse transport MFP. The first inverse MFP, G_1, is often referred to as the scattering power. As F_{L} depends only on the dynamic variables ω and E, the change of variable (7) and the condition (10) can be applied as before with the GS distribution. Again, this process splits F_{L} into a screened Rutherford PDF and a $q(u; E)$ surface, which now includes the effect of the energy losses within the CSDA model. Neither the memory storage nor the simulation time are significantly affected by this change. However, because the integral in (13) must be evaluated numerically for each ℓ, the computation time required to generate the table from which q will be interpolated during the execution of DPM does increase considerably. Fortunately, this needs to be done just once for a given material. It is worth noting that, unlike other schemes (Kawrakow and Bielajew 1998a), the method presented here is rigorously exact within the CSDA model, and the only approximation introduced involves the assumption of a screened Rutherford potential.

The importance of including energy losses in the multiple-scattering theory for large pathlengths is apparent from figure 2, which shows the difference in depth dose profiles for a 10 MeV electron beam in water when the more accurate Lewis approach is used instead of the GS scheme.

2.3. Electron transport mechanics

Since large pathlengths must be used to attain appreciable speed up of MC electron transport computations, the mechanism used to generate final phase space variables after a transport step plays a critical role in determining the accuracy of the model. As noted earlier, the efficacy of a given transport model can be evaluated by its faithfulness in reproducing the spatial and angular moments of the phase variables and the spatial and angular distributions, at the end of a given step. A comparison of transport mechanics methods has been performed by Larsen (1992) and by Kawrakow and Bielajew (1998a). They concluded that when the energy loss along a step is disregarded, PENELOPE's random hinge model (Fernández-Varea *et al* 1993) provides an excellent compromise between speed and accuracy, and is therefore well suited for a fast MC code. The random hinge transport method is described as follows. The pathlength s is split in two substeps of lengths

$$s_{\mathrm{A}} = \xi s \qquad \text{and} \qquad s_{\mathrm{B}} = s - s_{\mathrm{A}} \tag{15}$$

respectively, where ξ is a random number between 0 and 1. A first substep s_{A} is taken in the initial electron direction, after which the particle is deflected according to any multiple-scattering law which provides polar and azimuthal deflection angles Θ and Φ *determined over the entire step* s. A second substep is then taken over the remaining distance s_{B} in the new direction. For a particle directed along the z-axis starting at location $\vec{x} = 0$, provided that

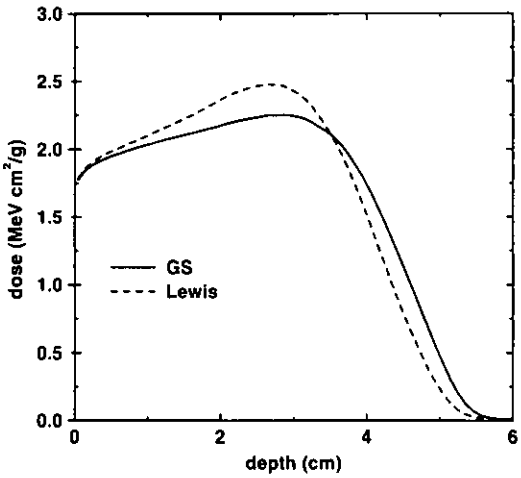


Figure 2. Depth dose profiles of a 10 MeV electron beam in water calculated showing the effect of substituting a model based on the GS theory (full curve) by another based on the Lewis' theory (broken curve). The pathlength was set to 1 cm.

the scattering law is correct and disregarding energy losses along the step, it can be shown that this method yields average values of the normalized penetration depth z/s and lateral displacement $(x^2 + y^2)/s^2$, which are correct to $O(s)$. Other moments preserved to this order accuracy include $(zv_z)/s$ (for $|v| = 1$), $(xv_x + yv_y)/s$, z^2/s^2 and $(x^2 + y^2)z/s^3$.

The inclusion of energy losses along s reduces the accuracy of the random hinge model. Indeed, Larsen's analysis of the spatial moments is valid only when the scattering power, G_1, and other inverse transport MFPs do not depend on s, or equivalently, on the energy E. For long steps s, these conditions will not be met and a modified formulation of random hinge mechanics must be employed. We begin by noting that exact Lewis moments (indicated by $\langle\cdot\rangle_L$) under the CSDA energy-loss model are given by

$$\frac{\langle z\rangle_L}{s} = \frac{1}{s}\int_0^s ds' \exp[-K_1(s')] \simeq 1 - \frac{1}{s}\int_0^s ds' K_1(s') \simeq 1 - \frac{1}{2}sG_1\left(\frac{s}{3}\right) \tag{16}$$

for the penetration depth and

$$\begin{aligned}\frac{\langle x^2 + y^2\rangle_L}{s^2} &= \frac{4}{3s^2}\int_0^s ds' \exp[-K_1(s')]\int_0^{s'} ds'' \exp[K_1(s'')]\left(1 - \exp[-K_2(s'')]\right) \\ &\simeq \frac{4}{3s^2}\int_0^s ds'\int_0^{s'} ds'' K_2(s'') \simeq \frac{2}{9}sG_2\left(\frac{s}{4}\right)\end{aligned} \tag{17}$$

for the lateral displacement. K_ℓ is defined as

$$K_\ell(s) \equiv \int_0^s ds' G_\ell(s') \simeq sG_\ell\left(\frac{s}{2}\right) \tag{18}$$

with K_1 called the 'scattering strength'. These relations use a first-order Taylor expansion for $G_\ell(s)$, and neglect terms containing products of two or more K_ℓs. This gives an upper bound for the Lewis lateral displacement and an acceptable approximation for values of s not extremely large and energies not too low. If energy loss is taken into account in sampling $\cos\theta$

and PENELOPE's mechanics (equation (15)) is used otherwise, the random hinge moments can be computed to be (Fernández-Varea *et al* 1993)

$$\frac{\langle z\rangle_{\mathrm{P}}}{s} = \frac{1+\langle\cos\theta\rangle}{2} = \frac{1+\exp[-K_1(s)]}{2} \simeq 1 - \frac{1}{2}sG_1\left(\frac{s}{2}\right) \tag{19}$$

and

$$\frac{\langle x^2+y^2\rangle_{\mathrm{P}}}{s^2} = \frac{1}{3}(1-\langle\cos^2\theta\rangle) = \frac{2}{9}(1-\exp[-K_2(s)]) \simeq \frac{2}{9}sG_2\left(\frac{s}{2}\right). \tag{20}$$

We therefore see that the effect of ignoring energy loss in the transport mechanics is equivalent to evaluating the $G_\ell(s)$s at the step mid-point rather than the correct distances of $s/3$ or $s/4$. As $G_\ell(s)$ increases as s increases (and E decreases), $\langle z\rangle_{\mathrm{P}}$ slightly underestimates the true value $\langle z\rangle_{\mathrm{L}}$, and $\langle x^2+y^2\rangle_{\mathrm{P}}$ substantially overestimates $\langle x^2+y^2\rangle_{\mathrm{L}}$. In physical terms, the PENELOPE model overestimates the scattering for very large pathlengths. Computations of these moments for 10 MeV electrons in water with a pathlength of 1 cm show that the PENELOPE model is in error by less than 0.1% for the penetration $\langle z\rangle$, but gives an excess lateral displacement of about 3%. These discrepancies increase with decreasing energy, with the lateral displacement error rising to approximately 10% at 4 MeV. Moreover, this deviation is systematic and compounds as the total pathlength travelled by the electron increases.

The above analysis suggests a modification of the random hinge model which preserves the Lewis moments by sampling uniformly in scattering strength $K_1^{(\mathrm{A})}$ rather than in distance s, i.e.

$$K_1^{(\mathrm{A})} = \xi K_1(s). \tag{21}$$

An electron is then transported until it 'accumulates' a scattering strength equal to $K_1^{(\mathrm{A})}$, where a deflection is imposed. The electron is then moved the pathlength required to exhaust the scattering strength $K_1(s)-K_1^{(A)}$. It can be shown that this non-uniform PDF for the first substep distance s_{A} yields correct values for the average penetration depth, lateral displacement and other spatial moments to first order in sG_1 and sG_2 when a linear approximation is adopted for $G_\ell(s)$. Perhaps more importantly, in addition to correcting for the scattering overestimation, this new transport mechanism also provides a basis for simulating scattering across material or density boundaries. Time saved in multi-voxel transport offsets by far the additional book-keeping required in calculating the K_1 accumulated over the steps. The details of the transport through voxels is presented in a subsequent section.

3. Discrete electron energy loss interactions

DPM employs what Berger (1963) has categorized as a class II mixed simulation scheme for energy losses. Hard interactions, i.e. those yielding energy loss above given cut-offs, are simulated discretely using an analogue (event-by-event) model. Soft events, which are much more frequent but result in energy transfer below the cut-offs, are modelled as contributing to a continuous deposition of energy throughout the transport step, and are accounted for in the CSDA approximation through the use of a restricted stopping power, as described later. In addition to resulting in energy loss to the primary electron, hard ionization events generate secondary electrons and hard bremsstrahlung collisions generate secondary photons. The energy loss of the primary and the phase state of the secondary particles is generated by sampling from the appropriate PDFs describing the processes.

3.1. Hard inelastic collisions

DPM uses the Møller DCS, σ_M, to treat inelastic collisions of electrons with atomic electrons. The Møller cross section, which was derived for collisions with free electrons at rest, is given by

$$\sigma_M(k) = \frac{2\pi e^4}{mv^2}\frac{Z}{Ek^2}\left[1 + \left(\frac{k}{1-k}\right)^2 - \frac{k}{1-k} + \left(\frac{\gamma-1}{\gamma}\right)^2\left(k^2 + \frac{k}{1-k}\right)\right] \tag{22}$$

where e is the charge of the electron, m its rest mass, v its speed, γ is the ratio of its total energy $E + mc^2$ to its rest energy, Z is the number of electrons in the target molecule and $k = W/E$ is the fraction of kinetic energy lost. Note that in the Møller formalism, the maximum allowed value of k is $1/2$ due to the indistinguishability of the projectile and target electrons.

The inverse MFP λ_M^{-1} for hard inelastic events (those above the cut-off W_M) in homogeneous media is easily derived by integrating (22)

$$\lambda_M^{-1} = \frac{2\pi e^4}{mv^2}\frac{ZN}{E}\left\{\frac{1-2k_M}{k_M(1-k_M)} + \left(\frac{\gamma-1}{\gamma}\right)^2\left(\frac{1}{2} - k_M\right) + \left[\left(\frac{\gamma-1}{\gamma}\right)^2 - 1\right]\ln\frac{1-k_M}{k_M}\right\} \tag{23}$$

where N is the number of molecules per unit volume and

$$k_M \equiv \frac{W_M}{E}. \tag{24}$$

In the limit that $mc^2 \ll E$ and $k_M \ll 1$, equation (23) can be approximated as

$$\lambda_M \simeq \frac{A}{Z\rho}\frac{mc^2 W_M}{N_A 2\pi e^4} \tag{25}$$

where A is the atomic weight of the species and N_A is Avogadro's number. By default, DPM sets $W_M = 200$ keV, as knock-on electrons with less than that energy have ranges much smaller than the minimum 1 mm voxel size. For water, (23) yields a value of λ_M roughly equal to 2 cm and practically independent of E, as shown by equation (25). The simple $Z\rho/A$ dependence on medium composition implicit in (23) will be exploited later in transport across voxel boundaries.

When a Møller interaction takes place, the fraction k of energy lost is sampled from the normalized PDF based on (22) by combining the rejection and composition methods (Salvat and Fernández-Varea 1992), and a knock-on electron is generated and its energy, direction and position stored for later transport. Since energy losses are usually much larger than the binding energies, the approximation that target electrons are initially free and at rest is appropriate. A knock-on electron will then have a kinetic energy equal to W and a direction of movement determined by the conservation of momentum. Naming θ_2 as the angle formed between this direction and the velocity of the incoming electron, it is found that

$$\cos\theta_2 = \sqrt{\frac{W(E+2mc^2)}{E(W+2mc^2)}}. \tag{26}$$

3.2. Hard bremsstrahlung interactions

The bremsstrahlung DCS for an electron impinging on a neutral atom with Z electrons to produce a photon with energy $W = kE$ can be written as

$$\sigma_B(k) = \frac{Z^2}{\beta^2 k} f(k) \tag{27}$$

where β is the electron velocity in units of the speed of light. Except for very high values of Z and low values of E, the leading term in (27) removes almost all the dependence of $\sigma_B(k)$ on E and Z, and the correction f is a smooth function of k. Seltzer and Berger (1985) have given a tabulation of $f(k)$ in terms of Z and E for selected Z values. For materials and energies typically seen in radiotherapy problems, the data contained in these tables can be roughly approximated by means of a linear function

$$f(k) = a(1 - bk) \tag{28}$$

with a and b being material- and energy-independent constants selected by performing a fit to the tabulated data. Inaccuracies in this approximation have little effect for most problems of interest, as the parameters a and b are weak functions of E and Z in the radiotherapy regime.

For compounds or mixtures, DPM relies on the additivity rule, replacing Z^2 in equation (27) with

$$Z_{eq}^2 \equiv \sum_i q_i Z_i^2 \tag{29}$$

where q_i and Z_i represent the stoichiometric index and the atomic number of the ith atom respectively. (For the sake of simplicity, the symbol Z^2, will be used throughout in place of Z_{eq}^2.)

For a given cut-off energy for bremsstrahlung production W_B, the inverse MFP resulting from this approximate DCS is

$$\lambda_B^{-1} = \frac{Z^2 n_m a}{\beta^2}\left(\ln\frac{1}{k_B} - b(1 - k_B)\right) \tag{30}$$

where $k_B = W_B/E$. In the limit $mc^2 \ll E$ and $k_B \ll 1$, equation (30) can be approximated as

$$\lambda_B \simeq \frac{A}{Z^2 \rho N_A a}\left(\ln\frac{E}{W_B} - b\right)^{-1} \tag{31}$$

which shows that λ_B has a mild variation with E at high energies. This fact, along with the linear scaling of λ_B with $Z^2\rho/A$ will be used to facilitate cross-voxel transport.

The analogue simulation of hard bremsstrahlung events, despite their infrequent occurrence, is necessary to accurately reproduce the fluctuations of the kinetic energy of impinging electrons. As the Møller DCS depends on the energy loss roughly as k^{-2} and the bremsstrahlung DCS as k^{-1} (equations (22) and (27) respectively), large energy losses are more likely to happen when the latter type of interaction occurs. As a result, a non-negligible fraction of incident electron energy straggling is caused by bremsstrahlung.

The random sampling of the PDF corresponding to the normalized $\sigma_B(k)$ can be performed using $f(k)$ in equation (27) as a rejection function. The angular deflection of the incoming electron is small and can be neglected and the scattering angle of the secondary photon is set equal to its mean value, approximately given by (Heitler 1954)

$$\langle\theta\rangle \simeq \frac{mc^2}{E + mc^2} \tag{32}$$

which is the approximation adopted in the original version of EGS.

4. Photon interactions

Photon transport is described following a conventional analogue MC treatment until the energy falls below some user-defined absorption energy. Three processes, photoelectric absorption, Compton scattering and pair production, are considered. The inverse MFPs for

these interactions are taken from those generated by the PENELOPE preprocessing program MATERIAL. In radiotherapy class problems, Compton scattering is the only significant dose delivery mechanism, and so approximations have been adopted for treating photoelectric and pair production interactions.

Photoelectric absorption, which is relevant only at very low energies and for high atomic numbers, is simulated by assuming that all the energy is locally deposited. DPM does not generate secondary electrons or relaxation radiation, so it is therefore convenient to set the electron and photon absorption energies above the highest absorption edge of the highest Z material in the problem. In most applications, DPM uses absorption thresholds of 50 keV for photons and 200 keV for electrons.

Pair production is important only at the high end of the energy range relevant to radiotherapy, and only for high atomic numbers, and so some very rough approximations are made. DPM assumes that for the first emerging particle, all kinetic energies are equally probable, and generates two electrons travelling in the same direction as the incident photon. Both particles are tracked as electrons, and one is randomly selected upon stopping to emit two annihilation photons travelling in randomly selected opposite directions. This approach disregards the differences in the cross sections and stopping powers between the created electron and positron (and the small possibility of in-flight annihilation of the positron), approximations which are justified by the relatively small impact of this effect in practical problems.

Compton interactions are assumed to involve free electrons at rest, and therefore binding effects (accounted for by means of the incoherent scattering function) and the Doppler broadening (Ribberfors 1975) of the energy of the scattered photon are ignored. These approximations are in general excellent for energies above 1 MeV, and are even more applicable to radiotherapy problems, where the only sources of low-energy photons are either contamination in the accelerator head and relatively rare hard bremsstrahlung interactions. DPM determines the energy of the scattered photon by sampling the Klein–Nishina DCS using the recipe contained in EGS4 (Nelson *et al* 1985). The recoil electron, which has direction and energy determined by the energy and momentum conservation laws, is stored in the secondary stack and simulated afterwards.

Photon histories terminate when the energy falls below a user-defined absorption energy or when they reach the geometry limits.

5. Transport across inhomogeneous voxel boundaries

As noted previously, a Monte Carlo electron transport algorithm sufficiently fast for clinical radiotherapy treatment planning will require the use of transport steps significantly greater than patient geometry voxel dimensions. This is problematic, as patient geometry (composition and density), which is typically inferred from CT data†, varies across almost every voxel boundary. Conventional MC programs are not capable of single-step transport over boundaries between differing media because the cross sections used in their multiple-scattering laws are medium dependent. Because DPM uses the medium-invariant $q(u)$ function to describe multiple scattering, however, transport across inhomogeneities is possible, as described below.

† Following the method of Knöös *et al* (1986), the program CTCREATE (Ma *et al* 1995) has been developed for the OMEGA BEAM (Rogers *et al* 1995) project, and is publicly available.

5.1. Electrons

The atomic number and density dependences of the physical models adopted by DPM have been examined in detail in the preceding sections. Here we describe how the particular forms of these quantities can be exploited to permit rapid simulation of electron transport in voxelized geometries.

In typical mixed class II electron transport MC models, electrons are started in an initial direction with an initial energy and transported in a series of steps until they exit the problem geometry or their energies fall below a user defined absorption cut-off. A transport step involves linear translation of the particle along its direction vector until a boundary is crossed, a hard collision takes place, or a multiple elastic scattering event is imposed. The details of how DPM determines when various events occur is presented here. Note that the computations described below of the distances travelled prior to the simulation of the different events are done in parallel as the electrons traverse the voxels.

- The distance to a Møller collision is sampled according to

$$t_M = -\lambda_M \ln \xi \tag{33}$$

where λ_M is the Møller MFP, equation (23), for some reference material (which will be assumed to be, without loss of generality, pure water) and ξ represents a random number uniformly distributed in (0, 1). A look-up table with values of λ_M on a grid of energies dense enough to allow accurate numerical interpolation is calculated beforehand and read from an input file during the initialization of DPM.

When an electron travels a distance t inside a voxel, t_M is decreased an amount Δt_M given by

$$\Delta t_M = t \frac{(Z\rho/A)_{vox}}{(Z\rho/A)_{water}}. \tag{34}$$

This is continued at each voxel until t_M drops to zero. A Møller interaction is then simulated, in which the energy lost by the incident electron is sampled from the Møller DCS of (22), and a knock-on electron is generated and placed in the secondary stack. Note that for a homogeneous medium made of water, the scattering event occurs when the distance t accumulated over voxels equals t_M.

- The distance to a bremsstrahlung collision is sampled according to

$$t_B = -\lambda_B \ln \xi \tag{35}$$

where λ_B is the bremsstrahlung MFP, equation (30), for the reference material, i.e. for water. Again, an interpolation table is generated beforehand for λ_B and read from an input file during the initialization of DPM.

When an electron travels a distance t inside a voxel, t_B is decremented by an amount equal to

$$\Delta t_B = t \frac{(Z^2\rho/A)_{vox}}{(Z^2\rho/A)_{water}} \tag{36}$$

and this process is repeated until t_B drops to zero. At that point a radiative event is simulated, in which a photon is generated with an energy sampled from the bremsstrahlung DCS of equation (27) and placed in the secondary stack. For a homogeneous medium made of water, the interaction takes place when the total distance s travelled across voxels equals t_B.

 295

- The total scattering strength K_1, given by equation (18) for $\ell = 1$, is obtained for water at the electron energy at the midpoint of the step. Values of K_1 as a function of energy and a preset pathlength s are precalculated and read by DPM during its initialization. The scattering strength prior to simulation of a multiple-scattering event is then sampled as

$$t_S = K_1^{(A)} \equiv \xi K_1 \tag{37}$$

in accordance with the corrected version of the PENELOPE transport mechanics described earlier (see equation (21)). It must be stressed that the units of t_S are *not* those of distance, but of scattering strength.

As each step t is taken inside a voxel, the scattering strength prior to multiple scattering, t_S, is decreased by an amount equal to

$$\Delta t_S = \int_0^t \mathrm{d}t' \, G_1^{(\mathrm{vox})}(t') \simeq \frac{t}{2}[G_1^{(\mathrm{vox})}(t'=0) + G_1^{(\mathrm{vox})}(t'=t)]. \tag{38}$$

Once t_S is exhausted, the angular deviation is sampled from the Lewis PDF, as described in previous sections, using the $q(u; E)$ surface corresponding also to water. After rotating through the scattering angle to determine the new electron direction, linear transport is resumed until a new quantity of scattering strength, given by

$$t_S = K_1^{(B)} \equiv K_1 - K_1^{(A)} \tag{39}$$

is spent. After the distance corresponding to t_S (as determined by summing the Δt_S incurred while stepping through each voxel) is traversed, the process is repeated, with a new total scattering strength K_1 determined by table look-up and a new t_S sampled according to (37). This procedure ensures that the actual pathlength is such that it produces the same mean angular deviation as over a predefined reference pathlength for water.

- Apart from discrete events, the continuous energy loss of the electrons is computed at each step. This is given by

$$\Delta E = \int_0^t \mathrm{d}t' \, S_r^{(\mathrm{vox})}(t') \tag{40}$$

where t is the distance traversed in a given voxel prior to a hard collision or exiting the voxel and $S_r^{(\mathrm{vox})}$ is the stopping power in the medium, 'restricted' to energy transfers below the Møller and bremsstrahlung production thresholds for the problem. For large t, $S_r^{(\mathrm{vox})}$ can vary over the step, and so the integral is approximated by first estimating the energy loss over t assuming that $S_r^{(\mathrm{vox})}$ is constant, and then averaging the stopping power over the step, as in

$$\Delta E = t \frac{S_r^{(\mathrm{vox})}|_{E_0} + S_r^{(\mathrm{vox})}|_{E_0 - t S_r^{(\mathrm{vox})}}}{2}. \tag{41}$$

Here E_0 is the electron kinetic energy at the beginning of the step and $E_0 - t S_r^{(\mathrm{vox})}$ is what the energy would be if the stopping power were constant. Values of $S_r^{(\mathrm{vox})}$ are precalculated for a dense grid of energies and read by DPM during its initialization.

It should be noted that, since hard inelastic MFPs depend on the energy, the exact sampling of the distance s to the next interaction of type 'i' is given by

$$\xi = \int_0^t \mathrm{d}t' \, \lambda_i^{-1}(t') \exp\left[- \int_0^{t'} \mathrm{d}t'' \lambda_i^{-1}(t'') \right]. \tag{42}$$

Equations (33) and (35) assume that $\lambda(t) \simeq$ constant across a voxel, which is computationally cheap, quite good for Møller collisions, and reasonably good for bremsstrahlung interactions. However, because of the possibility of large energy loss occurring in a hard collision, the energy

dependence of $\lambda(t)$ cannot be ignored when determining distances to additional interactions, and so t_M and t_B are recomputed individually whenever either type of hard collision takes place.

The electron history terminates when it leaves the CT geometry or when its energy falls below a user-defined absorption energy, E_{abs}, set by default to 200 keV that is the approximate energy at which the electron CSDA range equals 1 mm, a typical voxel size.

One important advantage of this algorithm over conventional MC electron transport schemes is that the scaling of the cross sections precludes expensive table look-ups when each new voxel is encountered. More significantly, the number of multiple-scattering events is dramatically reduced. In a conventional scheme, electrons are deflected not only when a multiple scatter step is traversed, but also at every boundary crossing and prior to the simulation of every hard inelastic collision. Thus DPM eliminates the majority of the computationally expensive samples from the MS distributions and rotations through the scattering angles.

5.2. Photons

Since photons undergo a limited number of interactions before they are locally absorbed, their transport is almost always treated in an analogue manner. This requires that the distance to collision be recomputed at every medium boundary, as, unlike the case of electrons, there are no simple scaling laws which can be applied. Since voxelized geometries can present frequent changes of material in short distances (relative to photon MFPs), this imposes a significant speed penalty.

To overcome this difficulty, DPM uses the δ-scattering method of Woodcock *et al* (1965), which avoids calculating intersections with the interfaces of all the visited voxels by exploiting the fact that the distribution of collision distances t contains the product of the probability of *not* colliding prior to t and the collision density. The method is implemented by first determining the energy-dependent minimum total MFP $\lambda_\gamma^{(min)}(E)$ in the entire geometry. A distance to the next interaction t is then sampled using $\lambda_\gamma^{(min)}(E)$, and the photon is transported through t, ignoring all boundary crossings. Next, the material of the current voxel is determined, which is simple and efficient for voxelized geometries. An interaction is simulated at t only with probability $\mathcal{P}$ equal to

$$\mathcal{P} = \frac{\lambda_\gamma^{(min)}}{\lambda_\gamma^{(vox)}} \tag{43}$$

where $\lambda_\gamma^{(vox)}$ represents the total MFP in the current voxel. If an interaction does not occur, the transport is continued. The quantity $1 - \mathcal{P}$, which is easily determined for every voxel, can be considered to be the probability of a 'fictitious' event occurring, in which no phase space change takes place. If an interaction does occur, its type is sampled according to the corresponding probabilities $\mathcal{P}_i$ (with 'i' representing Compton collision, photoelectric absorption or pair production), which are

$$\mathcal{P}_i = \frac{\lambda_\gamma^{(vox)}}{\lambda_i^{(vox)}}. \tag{44}$$

$\lambda_i^{(vox)}$ represents the MFP of the interaction 'i' in the current voxel. The Woodcock method will be efficient if there is only slight variation in λ throughout the geometry, and so time expended in stopping to analyse collisions which are then determined to be fictitious is less than the time saved by not stopping to recompute λ at each voxel boundary.

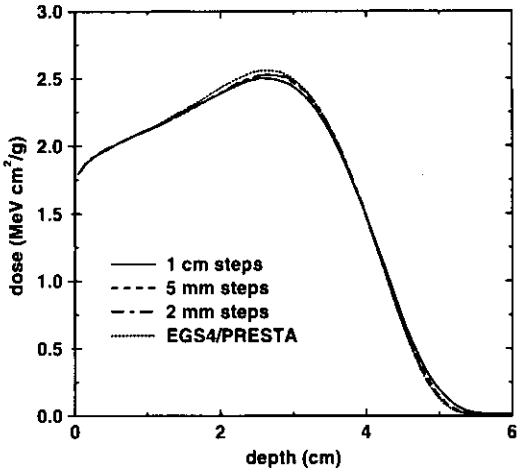


Figure 3. Depth dose produced by a 10 MeV electron pencil beam impinging normally on a semi-infinite water phantom using various reference step sizes.

6. Results

Results from simulations performed using DPM are compared here with results generated by EGS4 and PENELOPE. EGS4 has been extensively benchmarked against experimental data (Rogers and Bielajew 1989b, 1990) for radiotherapy problems, and is widely accepted as a standard. PENELOPE has likewise shown excellent agreement in a variety of comparisons with experimental and other Monte Carlo results (Baró *et al* 1995, Sempau *et al* 1997). Additionally, as DPM draws much of its physics data from PENELOPE's comprehensive and easily manipulated database, any discrepancies between DPM and PENELOPE should reflect differences in algorithms rather than differences in the underlying data or physical constants and models.

In order to fully exercise the approximations in DPM, a set of problems involving both homogeneous and multi-layered geometries and a wide range of materials (including several not typically seen in radiotherapy problems) has been simulated with all three codes.

6.1. Step size selection

The use of the condensed history method introduces an inherent error in Monte Carlo simulations, as elegantly characterized by Larsen (1992), who showed that it vanishes as the pathlength s tends to zero. But as efficient computation depends on taking long steps, ascertaining the longest multiple scattering step which preserves the accuracy of the simulation is of critical importance, and so is addressed by all conventional MC electron transport programs. Common treatments allow particles to advance until either some predetermined fraction of their initial kinetic energy is lost (e.g. PENELOPE when `C2` is active, EGS4 with the `ESTEPE` option, ETRAN, ITS, MCNP) or until some fixed pathlength has been travelled (e.g. PENELOPE when the option `HFPMAX` is active, EGS4 with the `SMAX` option). Another frequently used technique (e.g. PENELOPE when `C1` is active, EGS4 by default) limits s by fixing the mean angular deviation and calculating the step size accordingly.

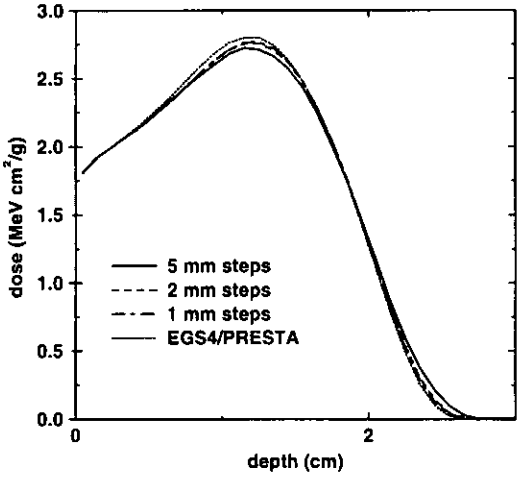


Figure 4. Depth dose produced by a 5 MeV electron pencil beam impinging normally on a semi-infinite water phantom using various reference step sizes.

In DPM, the step size issue arises even though $q(u)$ has been shown to be a function of u and E rather than s. The energy dependence of q is driven by the energy dependence of $s/\lambda(E)$, and since $\lambda(E)$ is fixed by the reference material, in order to use precomputed tabulated values of $q(u)$ in sampling for ω, s must be set in advance. The maximum value of s which preserves accuracy in a simulation can be determined by comparing simulation results for increasingly smaller step sizes. In figure 3, we show depth dose curves computed by DPM for one million 10 MeV electrons incident on a homogeneous phantom of 128^3 1 mm voxels, using $q(u)$ and K_1 calculated with step sizes ranging from 1 cm to 1 mm. There is little appreciable difference in the results for steps shorter than 5 mm. Note that because of the computation of energy loss in each voxel and the modelling of hard inelastic collisions, the computing time for a simulation is not directly proportional to the number of steps. The CPU usage for the simulations shown in this figure increase by only a factor of two while the step size decreased tenfold. At lower energies, there is more scatter for the same distance s, and the use of long steps maximizes the underlying condensed history error, as is seen in figure 4. In the first case ($s = 5$ mm), the fall of the DPM curve occurs just after completing the second step, reflecting the failure of the last one or two steps in reproducing the remaining part of the depth dose. This behaviour disappears when the pathlength is reduced to 1 or 2 mm. From these two sets of results, we determined that roughly 8–10 scattering events per history are necessary to reproduce depth dose profiles accurately, and that no further accuracy is attained by moving to steps of 1/20th of the range. By contrast, EGS4 requires the simulation of several multiple scattering events *in every voxel traversed by the electron.*

6.2. Homogeneous phantoms

Figures 5 to 9 show deposited depth dose curves for electron beams in semi-infinite phantoms made of different materials of interest in radiotherapy. The differences between EGS4, DPM and PENELOPE are equal to or less than 1.25% of the dose maximum in all cases and the statistical uncertainty of the curves presented are of the order of 0.2% of the dose maximum,

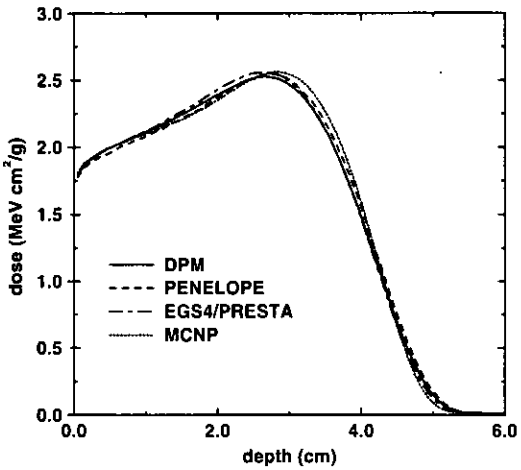


Figure 5. Depth dose produced by a 10 MeV electron pencil beam impinging normally on a semi-infinite water phantom.

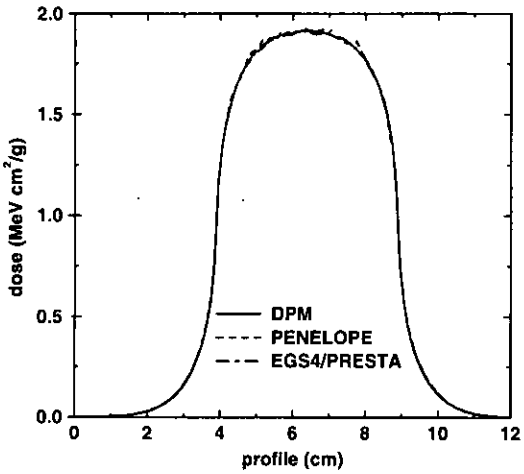


Figure 6. Dose integrated over planes perpendicular to the y-axis (that goes parallel to the water surface) produced by a 10 MeV electron beam impinging normally on a semi-infinite water phantom in a 5×5 cm^2 field.

and step sizes of 5 mm are used. Energy cut-offs of 200 keV for electrons and 50 keV for photons were used for all cases. In figure 5, the depth dose curve computed by MCNP is also included. It is interesting to note that the results from DPM generally lie inside the envelope of the results from the other programs.

6.3. Effects of hard collision physics approximations in DPM

The next set of problems was specifically designed to test the limits of the approximate scaling of the hard collision cross sections used in DPM. Recall that DPM samples the distance to a

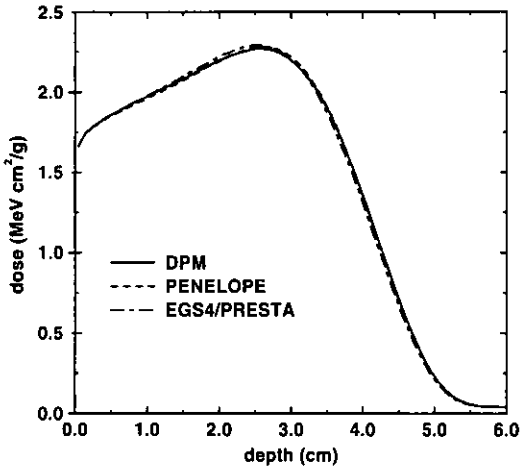


Figure 7. Depth dose for an 18 MeV normal pencil beam in a semi-infinite bone phantom.

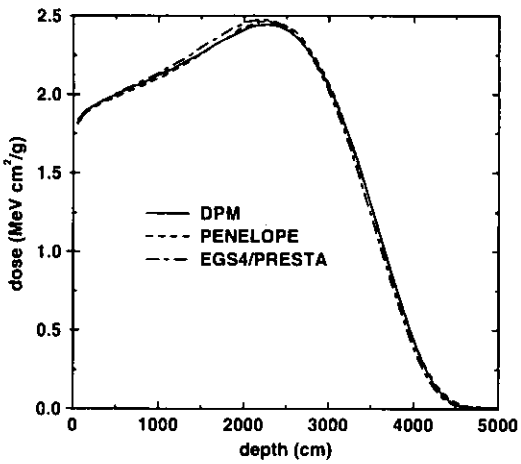


Figure 8. Depth dose for a 10 MeV normal pencil beam in an air phantom.

hard inelastic interaction as if the MFP were independent of energy along the path up to the interaction point. This approximation works well for Møller interactions due to the relatively slow variation of their MFP over the relevant energy range, as reflected in equation (25). Although still acceptable, it does not work equally well for bremsstrahlung, equation (31). Therefore a small error is introduced at those energies for which the bremsstrahlung contribution to the dose is not negligible. Since the slope of the bremsstrahlung MFP as a function of the electron energy is negative in the region of interest, DPM systematically underestimates this MFP. This effect is equivalent to an overestimation of the radiative stopping power, which will appear as an increase in the dose at shallow depths when it dominates over other sources of error.

In figures 10 and 11 depth dose curves in water for 15 and 20 MeV beams are represented to show the increase of the overestimation of the radiative stopping power as the energy of the

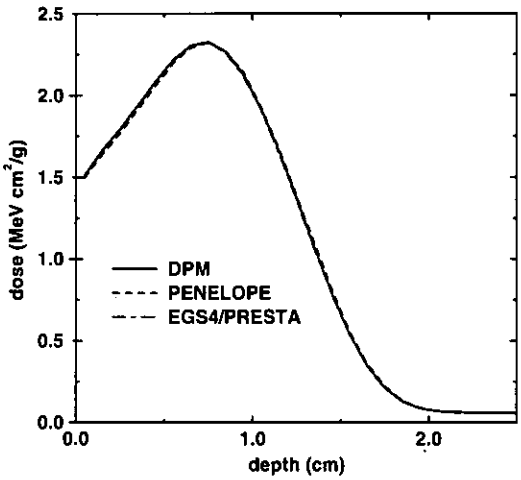


Figure 9. Depth dose for a 15 MeV normal pencil beam in a titanium phantom.

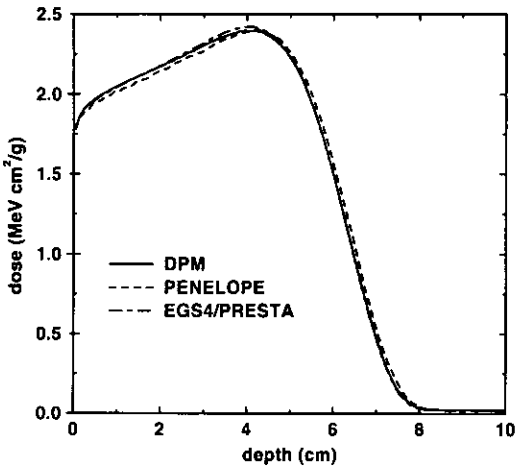


Figure 10. Depth dose for a 15 MeV pencil beam impinging on water.

beam increases. Despite the approximation on the bremsstrahlung cross section, the agreement is good and no correction for this effect is needed below 20 MeV.

Figure 12 shows the effect of these approximations in an extreme case, that is, for a very high-Z material and at the highest energy considered. With the current DPM model, discrepancies of up to 8% are seen between DPM and other MC programs. By switching to a method in which energy loss between collisions is accounted for in updating t_B, the difference between DPM and both PENELOPE and EGS4 can be reduced to 3–5%, as seen in the figure. However, as this introduces a computational overhead of close to 10% because of the frequency with which the cross sections must be computed and this effect is significant only for thick targets and very high Z, this correction is not retained in the basic DPM model.

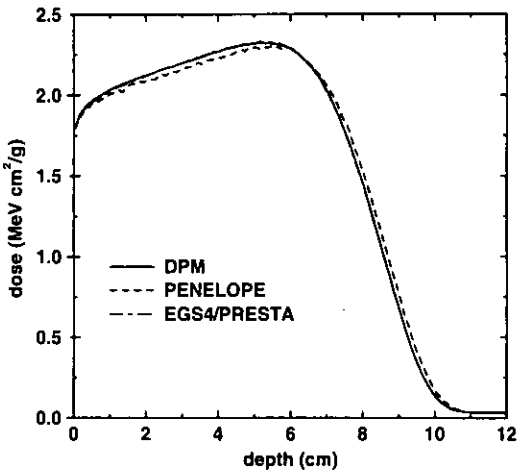


Figure 11. Depth dose for a 20 MeV pencil beam impinging on water.

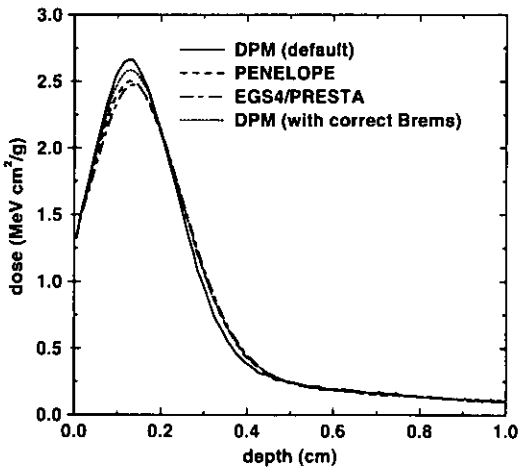


Figure 12. Depth dose produced by a 20 MeV pencil beam on a tungsten ($Z = 74$) phantom.

6.4. Inhomogeneities

Several multiple slab configurations were chosen to test DPM with inhomogeneous geometries, and results are presented in figures 13 to 15. Good agreement is found between the results from the three codes, with differences of the order of 1% of the dose maximum, reaching a maximum of 2% between PENELOPE and EGS4 or DPM in the higher-energy cases.

6.5. CT geometry

We present below results of dose computations using representative CT data to model density variation in a voxelized geometry. As the default PENELOPE package was designed to work with objects made of solid bodies with constant densities, it is unable to handle the density

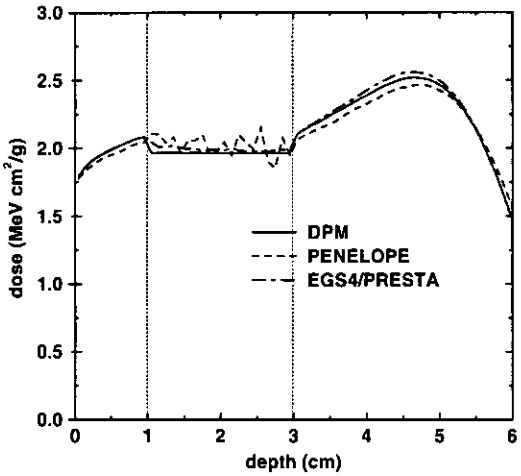


Figure 13. Depth dose in a water phantom with an air layer. The beam energy was 10 MeV.

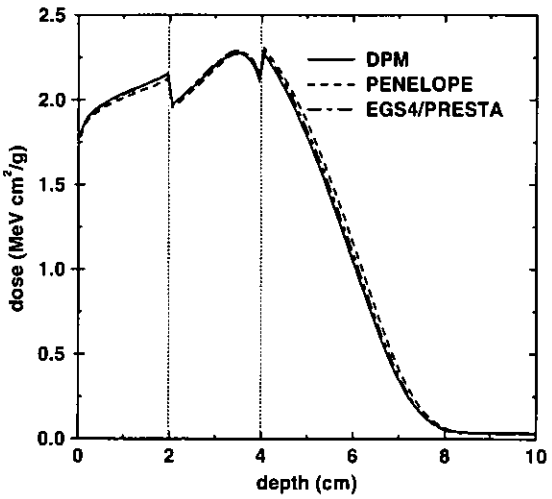


Figure 14. Depth dose in a water phantom with a bone layer. The beam energy was 18 MeV.

variations between neighbouring voxels present in the geometry of a real patient†. A utility for modelling density variations between regions does exist in the EGS4 system, and so this feature was exploited in generating simulation results for comparison with DPM.

In figure 16 a slice of the patient scan used for the current simulations is shown. Results from these calculations, which assumed a fictitious 16 MeV electron beam, are presented in figures 17 to 19. In order to facilitate comparisons of doses for individual voxels, a very large number of histories (10^9) were simulated. This number is orders of magnitude higher than that required for a routine treatment plan simulation, and gives a standard deviation of approximately 0.3% of the dose maximum for voxels of 1 mm on a side. The agreement

† A voxel based version of PENELOPE is currently being developed.

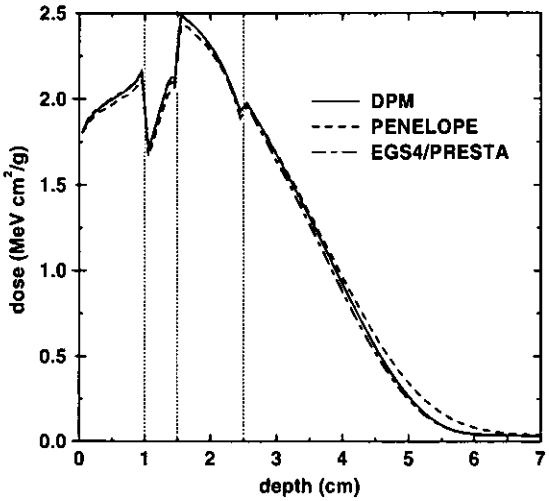


Figure 15. Depth dose in a multilayered water/titanium/bone/water geometry. The beam energy was 15 MeV.

between DPM and EGS4 is good, as expected from the previous results for homogeneous and multilayered geometries. The differences in the voxel doses are less than 3% of the overall maximum dose found in the geometry. If the differences are referred to the maximum dose in each figure, only the case with the lowest dose ($x = 28.5$ mm) significantly exceeds 3% of that maximum, ranging from 3 to 8%. Moreover, dose volume histograms (DVHs) were obtained for a specific target volume (or region of interest, ROI) of the same CT shown in figure 16. In figure 20 these DVHs are compared, showing an excellent agreement, with differences of the order of a few per cent.

7. Timing and efficiency

In table 1 we present the measured CPU times required to run 1 million histories on several of the test problems reported earlier. All runs were performed on an HP C3000 workstation, which is based on a 400 MHz HP PA8500 CPU. The program was compiled with the HP-UX f77 compiler and the recommended optimization switches `+O4 +E1 +E4 +E6 -K +U77`. The reference step size was chosen to be the largest such that the DPM results lay within 1% of the EGS4 results. This value was 0.5 cm in all cases except for 5 MeV on water, for which it was necessary to use a 0.2 cm step. Note that the 16 MeV CT case requires more time than the 20 MeV water case because the envelope of air and other less dense media under the beam give rise to deeper penetration of the source particles, requiring more computations of energy deposition and voxel crossings.

Table 2 presents results from a profiling study of DPM. Values are expressed in percentage of overall CPU time spent in various process. Two important conclusions can be drawn. First, the time taken in multiple scattering processes is quite small (~3%), implying that little further speed-up can be achieved in electron transport Monte Carlo simulations through manipulating step sizes. Second, a great deal of time (41%) is spent in voxel-to-voxel boundary crossings and CSDA energy deposition calculations, two unavoidable tasks of any algorithm based on mixed simulation of the inelastic interactions. Therefore, DPM can be considered

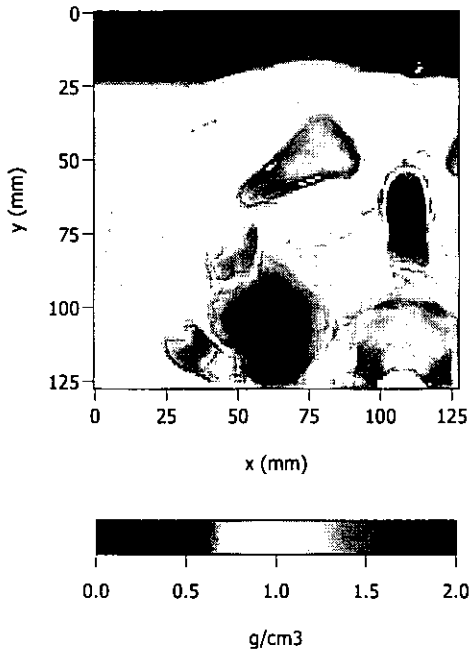


Figure 16. CT image representing a slice located 5.5 cm deep in the z direction, perpendicular to the paper. The x-axis goes from left to right and the y-axis points downwards. The 'universe' of the simulation consists in 128 × 128 × 128 cubic voxels of side 1 mm. A fictitious 16 MeV electron beam coming along the positive direction of the y-axis was defined, entering the universe through a 5 × 5 cm^2 square covering the range $(x, z) = (3.9 - 8.9, 3.9 - 8.9)$ cm. 10^9 histories were simulated to obtain the results presented in the next figures.

(This figure is in colour only in the electronic version, see `www.iop.org`)

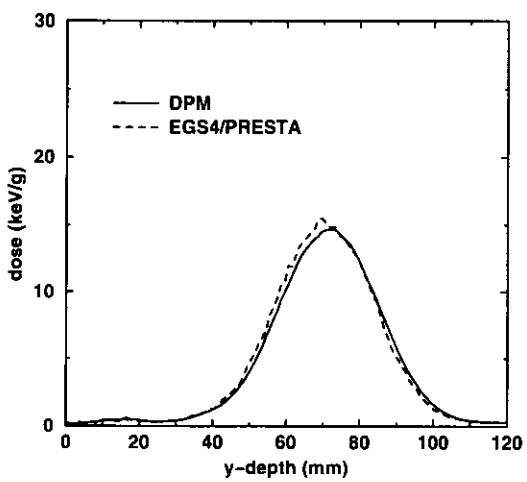


Figure 17. Dose along a line parallel to the y-axis of the CT slice represented in figure 16 at the value of $x = 28.5$ mm. Notice that the considered voxels lay outside the source field.

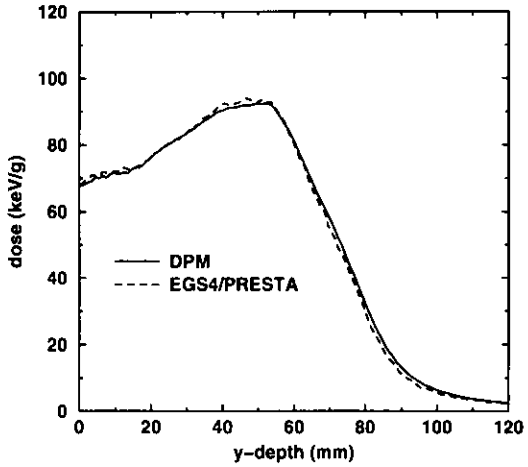


Figure 18. Dose along a line parallel to the y-axis of the CT slice represented in figure 16 at the value of $x = 60.5$ mm. Notice that the considered voxels lay directly under the source field.

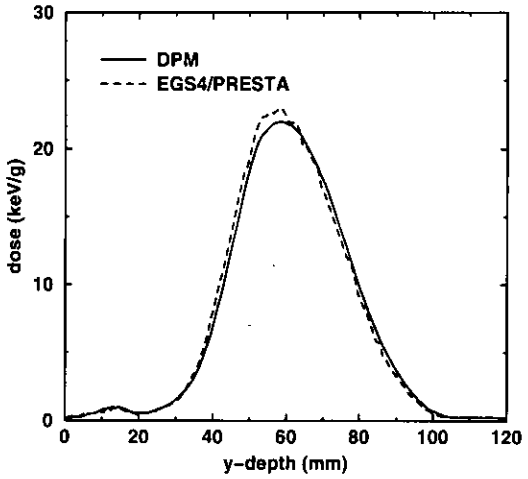


Figure 19. Dose along a line parallel to the y-axis of the CT slice represented in figure 16 at the value of $x = 94.5$ mm. The considered voxels lay outside the source field.

to exhibit close to the maximum achievable efficiency for condensed history CSDA MC codes.

8. Conclusion

A fast MC algorithm for the simulation of the dose deposited by electron–photon showers under radiotherapy conditions has been developed. DPM takes advantage of a new transport mechanics and an accurate multiple-scattering formalism independent of Z, permitting long simulation steps across media boundaries, significantly increasing the efficiency of the

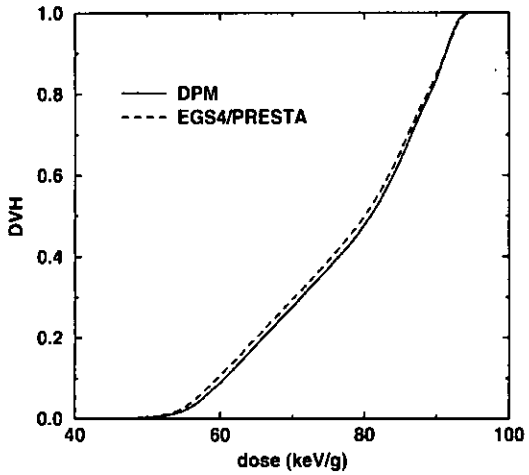


Figure 20. Integral DVHs obtained for the ROI defined by a cube of $2 \times 2 \times 2$ cm^3 centred at $x = 6.4$, $y = 5.9$ and $z = 5.5$ cm of the CT shown in figure 16.

Table 1. CPU time for one million histories.

Figure	Problem description	CPU time (s)
5	10 MeV pencil beam on water slab	169.4
6	10 MeV broad beam on water slab	181.4
4	5 MeV pencil beam on water slab	111.2
10	15 MeV pencil beam on water slab	250.6
11	20 MeV pencil beam on water slab	327.0
17–20	16 MeV broad beam on CT geometry	383.2

Table 2. Code profile. Use of CPU time in per cent for 16 MeV electrons incident on CT profile phantom.

86 e^- transport	=	
	55 transport through voxels	=
		14 geometry handling
		41 CSDA E loss and translation
	3 transport to collisions	
	3 sample scattering & rotations	
	25 data table look-ups	
2 photon transport		
1 I/O		
11 tallying		

computation without appreciably distorting the results. DPM has been shown to reproduce the dose distributions calculated with high-accuracy state-of-the-art general-purpose MC codes within an error of the order of 1.25% of the dose maximum, but with significant increase in computational efficiency. Dosimetric results accurate enough for electron beam radiotherapy applications can be generated in times of the order of 5 min on desktop workstations.

It has been pointed out (Bielajew 1994a, 1997) that present radiotherapy treatment planning systems, based on some type of analytic approximation to the solution of the transport equation will, some day, be replaced by systems based on the much more accurate and conceptually simpler MC methods. This work has shown that the use of these methods for both fast and accurate simulation of the transport of electrons in CT geometries is indeed feasible.

Acknowledgments

We would like to thank Drs Dick Fraass, Kwok Lam, Dan McShan and Randy Ten Haken, from the Radiation Oncology department of the University of Michigan, Dr Ed Larsen from the Department of Nuclear Engineering and Radiological Sciences, and Dr F Salvat, from the ECM Department of the Universitat de Barcelona, for fruitful discussions. One of the authors (JS) gratefully acknowledges the financial support of the Direcció General de Recerca de la Generalitat de Catalunya (Spain) grant no 1997BEAI400256 as well as the financial support from the Spanish Fondo de Investigación Sanitaria under contract no 98/0047-01. Financial support for this work has also been partially provided under the auspices of the US Department of Energy by the Lawrence Livermore National Laboratory under contract number W-7405-ENG-48, and by ADAC Laboratories (Milpitas, CA).

References

Andreo P 1991 Monte Carlo techniques in medical radiation physics *Phys. Med. Biol.* **36** 861–920

Baró J, Sempau J, Fernández-Varea J M and Salvat F 1995 PENELOPE: an algorithm for Monte Carlo simulation of the penetration and energy loss of electrons and positrons in matter *Nucl. Instrum. Methods* B **100** 31–46

Berger M J 1963 Monte Carlo calculation of the penetration and diffusion of fast charged particles *Methods Comput. Phys.* **1** 135–215

Bethe H A 1953 Molière's theory of multiple scattering *Phys. Rev.* **89** 1256–66

Bielajew A F 1994a Monte Carlo modelling in external electron-beam radiotherapy—why leave it to chance? *Proc. of the 11th Conf. on the Use of Computers in Radiotherapy* (Madison, WI: Medical Physics Publishing) pp 2–5

——1994b Plural and multiple small-angle scattering from a screened Rutherford cross section *Nucl. Instrum. Methods* B **86** 257–269

——1997 Monte Carlo dose calculation: why gamble with anything else? *Proc. of the First International Workshop on EGS4* (Tsukuba City, Japan: Laboratory for High Energy Physics) pp 310–23

Bielajew A F, Hirayama H, Nelson W R and Rogers D W O 1994 History, overview and recent improvements of EGS4 *National Research Council of Canada Report* PIRS-0436

Bielajew A F and Rogers D W O 1989 Variance-reduction techniques *Monte Carlo Transport of Electrons and Photons* ed T Jenkins *et al* (New York: Plenum) pp 407–19

Bielajew A F and Salvat F 2000 Improved electron transport mechanics in the PENELOPE Monte Carlo model *Nucl. Instrum. Methods* submitted

Bothe W 1921a Das allgemeine Fehlergesetz, die Schwankungen der Feldstärke in einem Dielektrikum und die Zerstreuung der α-Strahlen *Z. Phys.* **5** 63–9

——1921b Die Gültigkeitsgrenzen des Gaußschen Fehlergesetzes für unabhängige Elementarfehlerquellen *Z. Phys.* **4** 161–77

Briesmeister J F 1993 MCNP—a general Monte Carlo *N*-particle transport code *Los Alamos National Laboratory Report* LA-12625-M (Los Alamos, NM)

Cygler J, Battista J J, Scrimger J W, Mah E and Antolak J 1987 Electron dose distributions in experimental phantoms: a comparison with 2D pencil beam calculations *Phys. Med. Biol.* **32** 1073–86

Fernández-Varea J M, Mayol R, Baró J and Salvat F 1993 On the theory and simulation of multiple elastic scattering of electrons *Nucl. Instrum. Methods* **B73** 447–73

Goudsmit S A and Saunderson J L 1940a Multiple scattering of electrons *Phys. Rev.* **57** 24–9

——1940b Multiple scattering of electrons. II *Phys. Rev.* **58** 36–42

Halbleib J 1989 Structure and operation of the ITS code system *Monte Carlo Transport of Electrons and Photons* ed T Jenkins *et al* (New York: Plenum) pp 249–62

Halbleib J A, Kensek R P, Mehlhorn T A, Valdez G D, Seltzer S M and Berger M J 1992 ITS Version 3.0: the integrated TIGER series of coupled electron/photon Monte Carlo transport codes *Sandia Report* SAND91-1634

Hartmann-Siantar C L *et al* 1997 Lawrence Livermore National Laboratory's PEREGRINE project *Proc. of the 12th Conf. on the Use of Computers in Radiotherapy* (Madison, WI: Medical Physics Publishing) pp 19–22

Hartmann-Siantar C L, Chandler W P, Rathkopf J A, Svatos M M and White R M 1995 PEREGRINE: an all-particle Monte Carlo code for radiation therapy *Proc. Int. Conf. on Mathematics and Computations, Reactor Physics and Environmental Analyses* (La Grange Park, IL: American Nuclear Society Press) pp 857–65

Heitler W 1954 *The Quantum Theory of Radiation* (Oxford: Clarendon)

Kawrakow I and Bielajew A F 1998a On the condensed history technique for electron transport *Nucl. Instrum. Methods* B **142** 253–80

——1998b On the representation of electron multiple elastic-scattering distributions for Monte Carlo calculations *Nucl. Instrum. Methods* B **134** 325–36

Kawrakow I, Fippel M and Friedrich K 1996 3D electron dose calculation using a voxel based Monte Carlo algorithm *Med. Phys.* **23** 445–57

Keall P J and Hoban P W 1996 Super-Monte Carlo: a 3D electron beam dose calculation algorithm *Med. Phys.* **23** 2023–34

Knöös T, Nilsson M and Ahlgren L 1986 A method for conversion of Hounsfield number to electron density and prediction of macroscopic pair production cross-sections *Radiother. Oncol.* **5** 337–45

Larsen E W 1992 A theoretical derivation of the condensed history algorithm *Ann. Nucl. Energy* **19** 701–14

Lewis H W 1950 Multiple scattering in an infinite medium *Phys. Rev.* **78** 526–9

Ma C M, Mok E, Kapur A, Findley D, Brain S and Boyer A L 1999 Clinical implementation of a Monte Carlo treatment planning system *Med. Phys.* **26** 2133–43

Ma C M, Reckwerdt P, Holmes M, Rogers D W O and Geiser B 1995 DOSXYZ users manual *NRC Report* PIRS 509b

Mayol R and Salvat F 1997 Total and transport cross sections for elastic scattering of electrons by atoms *At. Data Nucl. Data Tables* **65** 55–154

Mohan R 1997 Why Monte Carlo? *Proceedings of the 12th Conference on the Use of Computers in Radiotherapy* (Madison, WI: Medical Physics Publishing) pp 16–18

Molière G Z 1947 Theorie der Streuung schneller geladener Teilchen. I. Einzelstreuung am abgeschirmten Coulomb-Field *Z. Naturforsch.* **2a** 133–45

——1948 Theorie der Streuung schneller geladener Teilchen. II. Mehrfach- und Vielfachstreuung *Z. Naturforsch.* **3a** 78–97

Nelson W R, Hirayama H and Rogers D W O 1985 The EGS4 code system *Stanford Linear Accelerator Center Report* SLAC–265

Neuenschwander H and Born E J 1992 A Macro Monte Carlo method for electron beam dose calculations *Phys. Med. Biol.* **37** 107–25

Neuenschwander H, Mackie T R and Reckwerdt P J 1995 MMC—A high-performance Monte Carlo code for electron beam treatment planning *Phys. Med. Biol.* **40** 543–74

Ribberfors R 1975 Relationship of the relativistic Compton cross section to the momentum distribution of bound electron states *Phys. Rev.* B **12** 2067–74

Rogers D W O and Bielajew A F 1989a A comparison of EGS and ETRAN *Monte Carlo Transport of Electrons and Photons* ed T Jenkins *et al* (New York: Plenum) pp 323–44

——1989b Experimental benchmarks of EGS *Monte Carlo Transport of Electrons and Photons* ed T Jenkins *et al* (New York: Plenum) pp 307–22

——1990 Monte Carlo techniques of electron and photon transport for radiation dosimetry *The Dosimetry of Ionizing Radiation* vol III, ed K Kase, B Bjärngard and F Attix (New York: Academic)

Rogers D W O, Ma C M, Ding G X and Walters B 1995 BEAM users manual *NRC Report* PIRS 509a

Salvat F and Fernández-Varea J M 1992 Semiempirical cross sections for the simulation of the energy loss of electrons and positrons in matter *Nucl. Instrum. Methods* B **63** 255–69

Salvat F, Fernández-Varea J M, Baró J and Sempau J 1996 PENELOPE, an algorithm and computer code for Monte Carlo simulation of electron-photon showers *Ciemat (Centro de Investigaciones Energéticas, Medioambientales y Tecnológicas) Technical Report* no 799

Seltzer S M 1989 An overview of ETRAN Monte Carlo methods *Monte Carlo Transport of Electrons and Photons* ed T Jenkins *et al* (New York: Plenum) pp 153–82

——1991 Electron-photon Monte Carlo calculations: the ETRAN code *Int. J. Appl. Radiat. Isotopes* **42** 917–41

Seltzer S M and Berger M J 1985 Bremsstrahlung spectra from electron interactions with screened atomic nuclei and orbital electrons *Nucl. Instrum. Methods Phys. Res.* B **12** 95–134

Sempau J, Acosta E, Baró J, Fernández-Varea J M and Salvat F 1997 An algorithm for Monte Carlo simulation of coupled electron–photon showers *Nucl. Instrum. Methods* B **132** 377–390

Wentzel G 1922 Zur theorie der streuung von β-strahlen *Ann. Phys.* **69** 335–68

Woodcock E, Murphy T, Hemmings P and Longworth S 1965 Techniques used in the GEM code for Monte Carlo neutronics calculations in reactors and other systems of complex geometry *Proc. Conf. on Applications of Computing Methods to Reactor Problems* Argonne National Laboratories Report ANL-7050

Photon beam relative dose validation of the DPM Monte Carlo code in lung-equivalent media

Indrin J. Chetty,[a] Paule M. Charland, Neelam Tyagi, Daniel L. McShan, and Benedick A. Fraass
The University of Michigan, Department of Radiation Oncology, Ann Arbor, Michigan 48109-0010

Alex F. Bielajew
The University of Michigan, Department of Nuclear Engineering, Ann Arbor, Michigan 48109-2104

(Received 8 November 2002; accepted for publication 2 January 2003; published 25 March 2003)

Validation experiments have been conducted using 6 and 15 MV photons in inhomogeneous (water/lung/water) media to benchmark the accuracy of the DPM Monte Carlo code for photon beam dose calculations. Small field sizes (down to 2×2 cm^2) and low-density media were chosen for this investigation because the intent was to test the DPM code under conditions where lateral electronic disequilibrium effects are emphasized. The treatment head components of a Varian 21EX linear accelerator, including the jaws (defining field sizes of 2×2, 3×3 and 10×10 cm^2), were simulated using the BEAMnrc code. The phase space files were integrated within the DPM code system, and central axis depth dose and profile calculations were compared against diode measurements in a homogeneous water phantom in order to validate the phase space. Results of the homogeneous phantom study indicated that the relative differences between DPM calculations and measurements were within ±1% (based on the rms deviation) for the depth dose curves; relative profile dose differences were on average within ±1%/1 mm. Depth dose and profile measurements were carried out using an ion-chamber and film, within an inhomogeneous phantom consisting of a 6 cm slab of lung-equivalent material embedded within solid water. For the inhomogeneous phantom experiment, DPM depth dose calculations were within ±1% (based on the rms deviation) of measurements; relative profile differences at depths within and beyond the lung were, on average, within ±2% in the inner and outer beam regions, and within 1–2 mm distance-to-agreement within the penumbral region. Relative point differences on the order of 2–3% were within the estimated experimental uncertainties. This work demonstrates that the DPM Monte Carlo code is capable of accurate photon beam dose calculations in situations where lateral electron disequilibrium effects are pronounced. © *2003 American Association of Physicists in Medicine.*
[DOI: 10.1118/1.1555671]

Key words: DPM Monte Carlo code, phase space, BEAMnrc, lateral electronic disequilibrium, ion chamber measurements

I. INTRODUCTION

Experimental validation of dose calculation models is an important step before the implementation of these algorithms in a clinical setting. Suggested dose accuracy for commissioning of treatment planning systems is typically 2%/2 mm in the high dose and penumbral regions, respectively, in homogeneous phantoms. These criteria are increased to 4%/4 mm in the presence of 3-D inhomogeneities,[1] where conventional dose algorithms do not offer explicit electron transport that is usually required to accurately characterize the perturbative effect of the inhomogeneity. The emergence of model-based dose calculation techniques, such as the convolution/superposition and the Monte Carlo methods, provide a more physics-based approach that has been found by many investigators to be more accurate than correction-based methods for calculating the dose in inhomogeneous media.[2–5] In particular, the Monte Carlo method is currently the only method that explicitly transports photons and electrons within a material and is therefore likely to provide more accurate results at material interfaces and within inhomogeneities. To date, researchers have conducted a wide range of experiments in both homogeneous and inhomogeneous geometries to validate user-specific Monte Carlo codes developed for clinical treatment planning.[6–21]

As physically realistic approaches become more practical for dose calculations, it becomes critical that these models be adequately validated against measurements. The increase in sophistication of dose algorithms also means that experimental validation should include complex geometries that aim to test the limits of the codes used. The focus of this work was to conduct experimental validation of the DPM (Dose Planning Method) Monte Carlo code for photon beam calculations in situations where lateral electron disequilibrium effects are emphasized, as observed, for example, when using small field sizes (2×2 cm^2), low-density media (lung-equivalent slabs) and high energies (15 MV photons). It was expected that such measurements would provide a stringent assessment of the transport physics employed within the DPM code.

Details of the electron/photon transport model used in

 0094-2405/2003/30(4)/563/11/$20.00

DPM are provided in the paper by Sempau *et al.*,[10] however, a few general comments on the transport physics are in order here. Electron transport within DPM uses a condensed history model that is based on a Class II "mixed" transport scheme for energy losses, with analog transport for large energy transfers, and the continuous slowing down approximation (CSDA) used for small energy losses.[10] DPM employs several features which make it optimal for radiotherapy class dose calculations. These include (a) the use of a step size independent multiple scattering theory based on the Kawrakow–Bielajew formalism,[10] (b) the use of a "random hinge" scheme for transporting charged particles from point-to-point in the medium, originally developed in the PENELOPE code[22] but modified within DPM to provide a basis for simulating scattering across material boundaries, (c) the use of large electron steps which affords the ability to traverse many voxels before sampling a multiple scattering angle, and (d) the use of Woodcock tracking (delta-scattering) to reduce the overheads associated with transporting photons across boundaries.

In this paper we present an investigation of the accuracy of DPM calculations versus measurements for photon beams incident upon a low-density composite water equivalent-lung equivalent phantom, for field sizes of 2×2, 3×3, and 10×10 cm^2. DPM calculations have initially been benchmarked in homogeneous geometries to validate the accuracy of the phase space simulation of the accelerator treatment head and ensure accurate modeling of the source. It should be noted that we have not included field sizes larger than 10×10 cm^2 in this study because the intent of this work was to investigate transport accuracy issues, which are best evaluated at small field sizes in low-density media. In describing the details of this work, the following topics will be addressed: Monte Carlo simulation of the linear accelerator treatment head using BEAMnrc, the experimental setup and measurement details, DPM simulation results and comparisons with measurements in both homogeneous and inhomogeneous phantoms.

II. SIMULATION OF THE LINEAR ACCELERATOR TREATMENT HEAD

A detailed phase space simulation of the components of a Varian 21EX linear accelerator (Varian Associates, Palo Alto, CA) was conducted using the usercode BEAMnrc[23] (based on EGSnrc[24] transport physics). The 21EX linac is an isocentric machine that produces two photon beam energies (6 and 15 MV) and five electron beams, from 6 to 20 MeV. BEAMnrc includes a comprehensive simulation geometry package that provides several component modules (CMs) with which to model various structures within the accelerator treatment head.[23] The specific CMs used for this study were the following: SLAB for the vacuum window, CONSTAK for the target and target housing, CONS3R for the primary collimator, FLATFILT for the 6 and 15 MV flattening filters, CHAMBER for the transmission chamber, MIRROR for the mirror, and JAWS for the secondary (x and y) collimators. The multi-leaf collimator was not included in this simulation

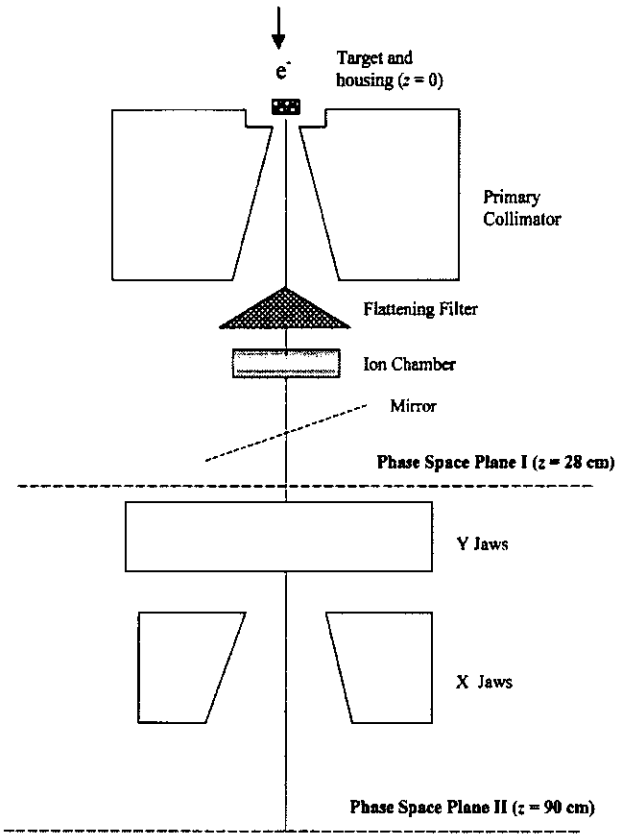


FIG. 1. BEAMnrc simulation geometry illustrating the treatment head components of the linear accelerator. The phase-space scored at plane I included only the patient-independent structures, while that scored at plane II included only the field-defining jaws.

and was retracted for all measurements performed in this study. Figure 1 illustrates the various components in the accelerator treatment head that were included in the simulation. In order to increase the efficiency of the simulation, a two-step process was utilized. In the first step, the phase space was tallied at a scoring plane (see phase space plane I, Fig. 1) located 28 cm downstream from the target and perpendicular to the beam central axis (CAX). This simulation included the "patient-independent" structures: the target, primary collimator, flattening filter, transmission chamber, and mirror. Specifically, the following parameters were scored for each history: x, y, u, v, *energy, latch, and weight*. The second step of the simulation process involved transport through the "patient-dependent" structures (the x and y secondary jaws). The jaws were set to define field sizes of 2×2, 3×3, and 10×10 cm^2 at the isocenter, and the phase space was scored (for each field size independently) at a plane 90 cm downstream from the target (see phase space plane II, Fig. 1). Note that the source particle parameters for the field-size-dependent simulation were contained in the phase space file acquired at plane I. All BEAMnrc phase space calculations in this work used default EGSnrc physics parameters.

The field-size-dependent phase space files were ported to the DPM code system for dose calculations within the phantom. A trial and error method, similar to that used by other

investigators,[11,17,23,25] was used to "calibrate" the incident electron-on-target energy for the initial phase space simulation. This involved adjusting the incident electron-on-target energy in the initial phase space simulations to provide the best fit between DPM calculations and measurements for central axis depth dose and profiles for a 10×10 cm^2 field size, 90 cm SSD, in water. For this study, the beam of electrons-on-target was modeled as a mono-energetic, parallel source of electrons with no angular spread. The "calibrated" electron-on-target energies were found to be 6.25 and 15.3 MeV for 6 and 15 MV photons, respectively.

III. PHANTOM MEASUREMENTS

A. Homogeneous phantom measurements

Central axis depth and profile doses were measured in the Scanditronix/Wellhöfer water scanning system using an SFD stereotactic diode (Scanditronix, Uppsala, Sweden) with a 2 mm active area diameter, and a 0.06 mm active volume thickness. The dimensions of this water phantom are $40\times40\times38$ cm^3. Measurements were conducted for 6 and 15 MV photons incident at 90 cm SSD, for field sizes of 2×2, 3×3 and 10×10 cm^2, defined at the isocenter. Profiles were measured at depths ranging from d_{max} (1.5 cm for 6 MV and 3.0 cm for 15 MV) to 30 cm. The photon diode was chosen for these measurements because of its superior spatial resolution which is necessary for accurately measuring small field profiles. Diodes have been used by other investigators[7] for small field measurements because of their enhanced spatial resolution.

B. Inhomogeneous phantom measurements

The inhomogeneous phantom consisted of slabs of solid water (Gammex RMI, Middleton WI) with dimensions of 30×30 cm^2 and varying thicknesses; the solid water material, from depths of 4 cm to 10 cm, was replaced with a 6 cm thick lung-equivalent full slab phantom of density 0.3 g/cm^3 (Gammex RMI, Middleton WI). It should be noted that the 6 cm lung-equivalent slab was constructed by re-arranging smaller pieces of lung-equivalent material of varying dimensions (varying in width, length, and thickness) to form the slab. The experimental setup for measurements within the inhomogeneous phantom is illustrated in Fig. 2. For the central axis depth dose measurements, an IC-10 (Scanditronix, Uppsala, Sweden) cylindrical ion chamber ionization chamber, with an air cavity volume of 0.13 cm^3 and a 3 mm inner radius, was inserted at the following depths within the phantom: 1.0, d_{max}, 2.0, 3.0, 5.0, 7.0, 9.0, 11.0, 13.0, 15.0, and 20.0 cm. Charge (in nC) was collected with a PRM Model SH-1 (Precision Radiation Measurements, Tennessee) electrometer operated at a 300 V bias. A set of 3 readings, with exposures of 100 MU per reading (corresponding to the calibration dose of 80 cGy at 10 cm depth, 90 cm SSD, 10×10 cm^2 in water), was acquired at each point. The effective point of measurement for the cylindrical ion chamber was taken into consideration by shifting the measured depth dose curve 1.8 mm for the 6 MV beam and 2.0 mm for the 15 MV beam, as recommended in the Scanditronix/Wellhöfer scanning system manual (Scanditronix, Uppsala, Sweden) and also used by others.[7]

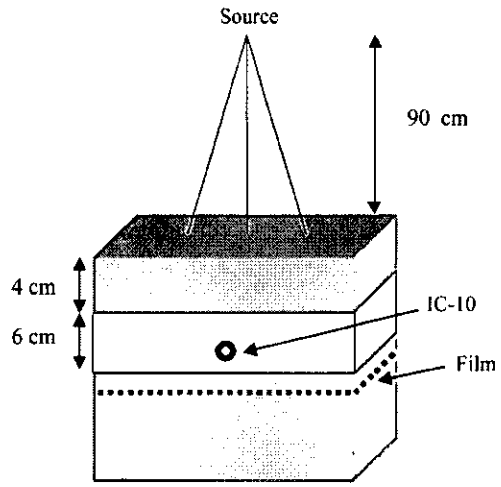


FIG. 2. Experimental geometry for the inhomogeneity measurements. The lung-equivalent slab extends from 4–10 cm within the solid water. An IC-10 ion chamber was inserted at various depths, within the solid water and lung, to measure the central axis depth dose. Film was inserted between slabs of solid water to measure profile doses.

Profiles along the central x-axis were measured with the IC-10 chamber at depths of 8 cm and 12 cm. This was accomplished by translating and repositioning the phantom so as to increment the chamber off-axis, by 0.5 cm in the high dose region of the profile and by 0.2 cm in the profile penumbral (high dose gradient) region. A graph paper with millimeter spacing was used to conduct the alignment and translation during this particular experiment. Profiles within the inhomogeneous phantom were also measured using Kodak Ready-Pack Extended Dose Range (EDR) film (Eastman Kodak Co., Rochester, NY). A recent study has shown that Kodak EDR film is dosimetrically comparable to the widely used Kodak XV film over a much larger range of doses.[26] In this experiment, films were sandwiched in the inhomogeneous phantom at depths of 8 cm and 12 cm. Each film pack was pin-pricked to avoid unwanted air in the envelope. Irradiated films were processed using a Kodak X_Omat-3000 RA, automatic film processor with a 90 second processing time. Processed films were converted to digitized images using the Lumiscan Model 100 (Lumisys, Sunnyvale, CA) laser digitizer with a 0.42 mm spot size and 0.45 mm pixel size. Profile images were analyzed with the Fuji film analysis software, ScienceLab 98—Image Gauge (Fuji Photo film Co, Ltd., Itasca, IL). The conversion from optical density to dose was carried out using an H&D or sensitometric curve for Kodak EDR film. Sensitometric curves were generated in solid water for 6 and 15 MV photons at a depth of 5 cm for a 10×10 cm^2 at 90 cm SSD, by exposing films (perpendicularly to the central axis) to varying doses from 0 to 400 cGy (in 12 equal dose increments), within the linear region of the sensitometric curve for Kodak EDR film. Optical densities for films irradiated for sensitometric analysis were read out using a Digital Densitometer II (Sun Nuclear Corporation,

Melbourne, FL) optical densitometer. Although the sensitometric curves were measured at a depth different than that of the profiles, the variation in sensitometric response with depth for the field sizes studied in this work was found to be negligible.[26]

C. Uncertainty estimates

For the diode depth dose and profile measurements in water, the uncertainty was estimated to be less than 0.5 mm in depth, relative to the surface. Positioning of the IC-10 ion chamber at a given depth was estimated to be within 1 mm, relative to the center of the chamber. Precision of the IC-10 chamber based on the reproducibility of at least three readings per point was found to be within $\pm 0.3\%$ (1σ) for all points measured. Measurements were taken at the beginning of the session and repeated a few hours later, at the end, to estimate drifts in the output of the accelerator. These differences were within $\pm 0.3\%$. The uncertainty associated with translation and alignment of the ion chamber with the crosshairs during profile measurements in the inhomogeneous phantom was estimated to be within 1 mm (2σ). For film measurements, uncertainties were estimated by identically irradiating and processing a set of eight films; the variation in optical density was found to be approximately $\pm 2\%$ (1σ).

IV. MONTE CARLO CALCULATIONS

A. Phase space source description

DPM calculations were conducted within a voxelized phantom using, for the input source description, the phase space files (at phase space plane II, Fig. 1) generated with BEAMnrc for 6 and 15 MV photons at 90 cm SSD in air, for 2×2, 3×3, and $10\times10\ \text{cm}^2$ field sizes as specified at the isocenter. Each record in the phase-space files contained the following parameters for each particle, in binary format: x, y, u, v, *energy, latch, and weight*. X and y represent the particle's position in the phase space plane (at a fixed z-location), u and v, the particle's direction cosine vectors along the x and y axes, respectively. The *latch* variable, in addition to storing the particle's creation/interaction history (dependent upon the latch type), also contains a bit to distinguish whether the particle is an electron or a photon.[23] A sub-routine was written to read the phase space parameters directly within the DPM code. Included within this routine is a calculation to determine w (the direction cosine vector along the z-axis) for each particle, based upon the method employed in DOSXYZnrc[16]—this calculation uses the identity $u^2+v^2+w^2=1$.

B. Monte Carlo uncertainty estimates

Both the DPM and BEAMnrc Monte Carlo codes utilize the history-by-history method for estimating the standard deviation, $S_{\bar{x}}$. This method has been described by Sempau *et al.*,[21] Briesmeister,[27] Walters *et al.*,[28] and Andreo.[2] $S_{\bar{x}}$ is calculated using the equation:

$$s_{\bar{x}}=\sqrt{\frac{1}{N-1}\left(\frac{\Sigma_{i=1}^{N}X_i^2}{N}-\left(\frac{\Sigma_{i=1}^{N}X_i}{N}\right)^2\right)}, \quad (1)$$

where N is the number of histories and X_i the quantity of interest (such as dose) scored in independent history i. The advantages of the history-by-history statistical estimator over the batch method are detailed in the paper by Walters *et al.*[28]

BEAMnrc calculations for the patient-independent structures (i.e., phase space acquired at plane I in Fig. 1) contained approximately 100 million histories in the phase space files for both 6 and 15 MV photons; the 1σ statistics in photon fluence was, on average, less than 0.5% for these runs. For calculations including the secondary jaws (i.e., phase space acquired at plane II in Fig. 1), the number of phase space records ranged from 2 million for the $2\times2\ \text{cm}^2$ field to 11 million for the $10\times10\ \text{cm}^2$ field, resulting in average statistics (1σ in photon fluence) ranging from 0.5–1.0%. Sempau *et al.*[21] have coined the term "latent uncertainty" to describe the uncertainty due to statistical fluctuations in the phase space; this is distinct from the uncertainty due to the random nature of the transport in phantom. As pointed out by Walters *et al.*,[28] the statistical uncertainty in calculated dose will approach the finite, latent uncertainty associated with the phase space, regardless of the number of times the phase space is sampled. In this work, the latent uncertainty was estimated by summing the uncertainties from phase space planes I and II in quadrature. In order to estimate the total uncertainty in the calculated dose in phantom, we have quadrature summed the uncertainties from the DPM phantom calculations (due to the random fluctuations in phantom) with the inherent (latent) uncertainty of the phase space (see Table I). These uncertainties fall roughly in the range from 0.5–1.5% (see Table I), for all points in this study.

C. DPM physics and scoring parameters

Calculations using the DPM Monte Carlo code were performed using a simulated cubic water phantom (with a side of 40 cm) for the homogenous geometry and a composite water-lung–water-equivalent phantom for the heterogeneous calculations. A scoring voxel with dimensions $2\times2\times2\ \text{mm}^3$ was used for most calculations; this was reduced to 1 mm in the scoring axis to obtain finer resolution for the smaller field (2×2, and $3\times3\ \text{cm}^2$) profiles. All calculations were conducted in water for a single, AP beam, perpendicularly incident on the phantom at 90 cm SSD, to mimic the measurement geometry. No differentiation was made between water and solid water in the input files for calculations in the lung-equivalent phantom; studies[26,29] have shown that relative depth dose differences between solid water and water are small (<1.0%), confirming that the comparison between measurements in water and Monte Carlo calculation solid water is unbiased by minor differences in the electron densities between these two materials. The lung-equivalent material was modeled as a slab with a uniform physical density of $0.3\ \text{g/cm}^3$ and an atomic composition specified by the

TABLE I. Quantitative analysis of the DPM calculated and measured depth dose curves illustrated in Figs. 4 and 7. Specifically shown are the rms deviations and the maximum point differences, the DPM (1σ) uncertainty range and the total (1σ) MC uncertainty range. The total MC uncertainty was estimated by summing the latent phase space and DPM uncertainties in quadrature.

Fig. No.	Description of experiment	rms % deviation (Maximum point difference)	DPM (1σ) uncertainty range (%)	Estimated total (1σ) MC uncertainty range (%)
4(a) (upper)	6 MV photons, 2×2 cm^2, CAX depth dose in water	0.86 (2.54)	0.39–0.78	0.6–1.4
4(a) (middle)	6 MV photons, 3×3 cm^2, CAX depth dose in water	0.75 (1.68)	0.36–0.74	0.6–1.3
4(a) (lower)	6 MV photons, 10×10 cm^2, CAX depth dose in water	0.55 (−1.53)	0.75–1.05	1.0–1.6
4(b) (upper)	15 MV photons, 2×2 cm^2, CAX depth dose in water	0.77 (−2.90)	0.07–0.14	0.5–1.1
4(b) (middle)	15 MV photons, 3×3 cm^2, CAX depth dose in water	0.87 (−2.14)	0.16–0.32	0.5–1.2
4(c) (lower)	15 MV photons, 10×10 cm^2, CAX depth dose in water	0.67 (−2.64)	0.34–0.64	0.7–1.4
7(a) (upper)	6 MV photons, 2×2 cm^2, CAX depth dose in water/lung/water	0.81 (−2.36)	0.08–0.16	0.5–1.1
7(a) (middle)	6 MV photons, 3×3 cm^2, CAX depth dose in water/lung/water	0.56 (−1.12)	0.12–0.22	0.5–1.1
7(a) (lower)	6 MV photons, 10×10 cm^2, CAX depth dose in water/lung/water	0.72 (−1.88)	0.34–0.60	0.7–1.4
7(b) (upper)	15 MV photons, 2×2 cm^2, CAX depth dose in water/lung/water	1.01 (−2.64)	0.07–0.11	0.5–1.1
7(b) (middle)	15 MV photons, 3×3 cm^2, CAX depth dose in water/lung/water	1.00 (−2.65)	0.10–0.16	0.5–1.1
7(b) (lower)	15 MV photons, 10×10 cm^2, CAX depth dose in water/lung/water	0.81 (−1.0)	0.25–0.37	0.7–1.3

manufacturer (Gammex, RMI, Middleton, WI). Measurements of the density of the lung-equivalent slabs, based on CT-scans and a physical measurement of the mass and volume, agreed with the value of 0.3 g/cm^3, as provided by the manufacturer. All DPM calculations were performed using a 1 mm step size, and low energy electron and photon cutoffs of 200 keV and 50 keV, respectively.

DPM is implemented on a linux-based platform consisting of multiple, 1.4 GHz (AMD Athlone) processors that are configured to process calculations in parallel. The shared cluster is owned and maintained by the University of Michigan Center for Advanced Computing. For this study, the number of processors for each calculation was dependent upon the availability of free processors, but typically ranged from 2 to 25. The time to process one hundred million histories on a single processor was approximately 0.9 h and 1.2 h for the 6 and 15 MV photon beams, respectively. The reduction in computing time was found to scale almost linearly with the number of processors with some minimal overhead associated with file I/O.

V. SIMULATION RESULTS AND DISCUSSION

A. Homogeneous phantom benchmarks

Figure 3 shows the central axis, normalized, bremsstrahlung spectra, differential in energy, for the 6 and 15 MV photon beams. The spectra were reconstructed, using BEAMDP, from the "open beam" phase space scored at phase space plane I in Fig. 1. Electron-on-target energies were 6.25 and 15.3 MeV for the 6 and 15 MV photons, respectively. From these figures, the average photon energies were computed to be 1.7 MeV (6 MV) and 3.62 MeV (15 MV). Figures 4(a) and 4(b) illustrate the relative central axis doses as a function of depth in water for the 6 and 15 MV photon beams respectively. DPM calculations are depicted with open markers and diode measurements are shown in the solid lines. Each figure depicts three sets of curves, corresponding to the square field sizes at the isocenter, 2×2, 3×3, and 10×10 cm^2, tested in this study. The curves have all been normalized at a depth of 10 cm; the 2×2 and 3×3 cm^2 fields include additional scaling factors of 0.50 and 0.75, respectively, for ease of illustration. The root-mean-square

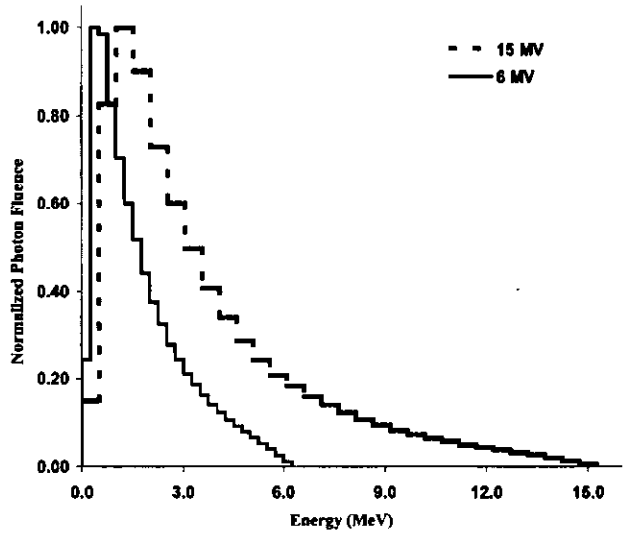


FIG. 3. Normalized photon fluence as a function of energy, along the central axis, for 6 and 15 MV photon beams. These bremsstrahlung spectra were reconstructed from phase space simulation of the linear accelerator treatment head using BEAMnrc. The electron-on-target energies were 6.25 MeV and 15.3 MeV for the 6 and 15 MV photon beams, respectively.

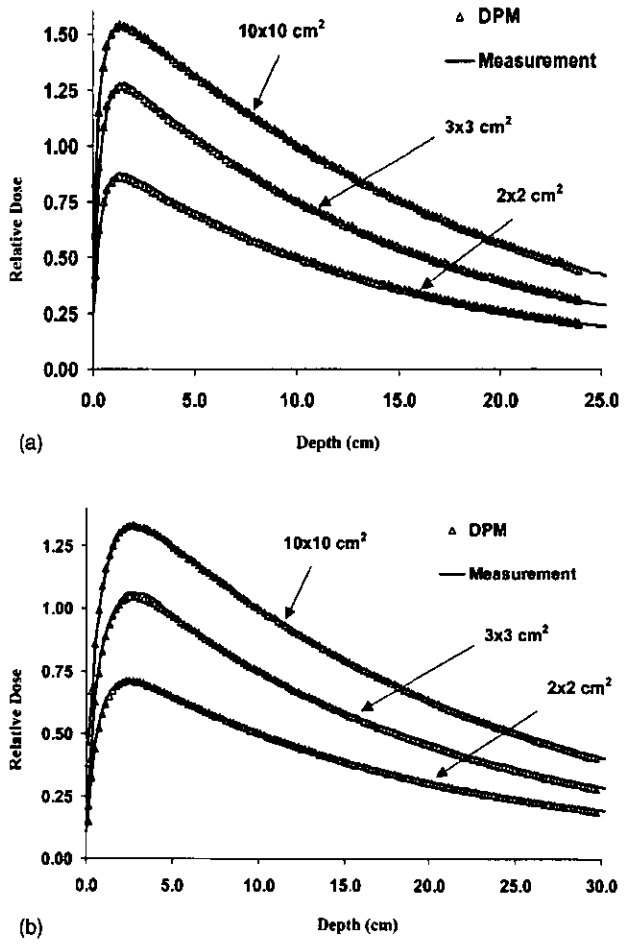


FIG. 4. Relative central axis depth dose for (a) 6 MV and (b) 15 MV photons in a water phantom. DPM calculations are shown with markers and diode measurements are shown in the solid lines. Curves are illustrated for field sizes of 2×2, 3×3, and 10×10 cm². Depth dose curves have been normalized to the point dose at a depth of 10 cm for the respective field size. The 2×2 and 3×3 cm² curves have been scaled using factors of 0.5 and 0.75, respectively, for an illustration on the same graph.

(rms) values of the percentage differences between calculations and measurements, as well as the maximum differences, are presented in Table I for all depth dose curves shown in Fig. 4. The percent differences were calculated using the relation: (calculated value−measured value)×100/maximum measured point value. Also included in Table I are the DPM (1σ) uncertainties for the in-phantom calculations, as well as the total Monte Carlo uncertainty, at the 1σ level. The total uncertainty was evaluated as a quadrature sum of the approximate latent uncertainty in the phase space file and the DPM uncertainty, due only to random fluctuations in phantom, as explained in Sec. IV A. As seen in Table I, DPM calculations are in within 1% agreement with measurements (based on rms % deviation) for all square field depth dose curves presented in Figs. 4(a) and 4(b). The rms % deviations are also within the estimated total Monte Carlo uncertainty ranges provided in Table I. Despite the good average agreement, we find maximum point differences in the range from −1.5% to −3% for the square field depth dose comparisons. The maximum point differences were all found to occur in the buildup region, where we estimate the experimental uncertainty to be on the order of 0.5 mm, or 5–10% per mm. Therefore, we attribute some part of these differences to the setup uncertainty. However, we also note in Table I that the maximum point differences for the 15 MV beam for the depth dose curves in Fig. 4(b) were all negative, indicating that the DPM calculated dose is less than measurements in the buildup region. This may lead one to suspect that there is potentially a systematic error that has not been accounted for. In fact, this may indeed be the case. One of the limitations of this study is that we have not thoroughly investigated buildup dose issues. It is well known that accurate measurement of dose (within ±3% uncertainty) requires an accurate setup as well as the use of extrapolation or parallel plane chambers. However, it is equally likely that the Monte Carlo calculations do not accurately (within ±3%) predict the dose in the buildup region.[7,30] For example, Hartmann Siantar *et al.*[7] have shown that their source calculations, conducted with PEREGRINE and based on phase space calculations from BEAM, required additional electrons to agree with measurements in the buildup region for field sizes ranging from 2×2–38×38 cm². Although the effect of contaminant electrons at small field sizes is small, it is nevertheless a potential source of the point dose disagreements observed in the buildup region. We are currently conducting a study of accurate dose measurements in the buildup region;[31] this data will be used to benchmark Monte Carlo calculated buildup doses over a range of clinically relevant field sizes.

Figures 5(a), 5(b), and 5(c) illustrate the 6 MV central axis profile doses along the x-axis in water for the 2×2, 3×3, and 10×10 cm² fields, respectively. DPM calculations are shown in open markers and diode measurements are shown in solid lines. Each figure depicts three curves, corresponding to profiles at depths of 1.5 cm (d_{max}), 10 cm, and

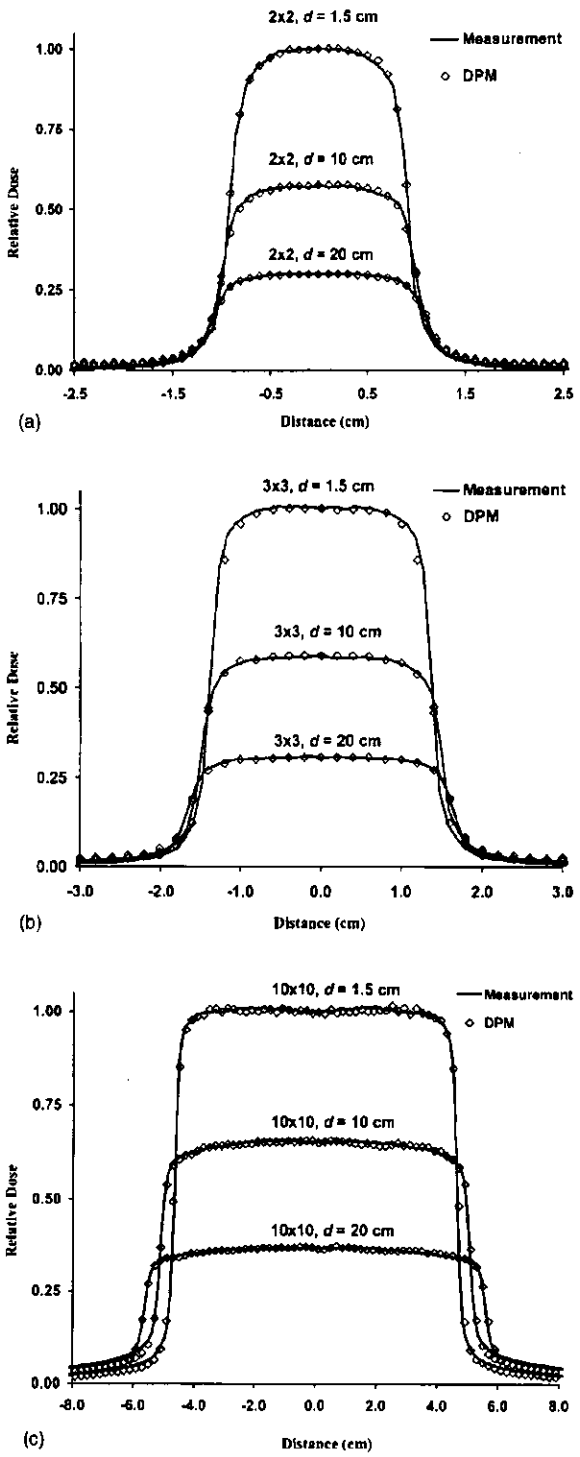


FIG. 5. Relative central axis profile doses for 6 MV photons in a water phantom. DPM calculations are shown with markers and diode measurements are shown in the solid lines. Profiles are illustrated for (a) 2×2 cm^2, (b) 3×3 cm^2, and (c) 10×10 cm^2 field sizes, at depths of d_{max} (1.5 cm), 10, and 20 cm. All curves have been normalized to the maximum central axis point dose.

20 cm. All curves have been normalized to the maximum point dose along the central axis. The 1σ total Monte Carlo uncertainties in these calculations ranged from less than 1% for the 2×2, and 3×3 cm^2 fields to less than 2% for the 10×10 cm^2 field size. Analyses of the differences between calculations and measurements for the profiles illustrated in Fig. 5 reveals that the average agreement is within (a) 1% in the inner beam region (dose>90%), (b) 1 mm distance-to-agreement in the penumbral region (20%<dose<80%), and (c) 2% in the outer beam region (dose<20%). These differences are within the acceptability criteria specified by the AAPM Task Group No. 53[32] for profile doses in homogeneous media. These suggested criteria are 2% (of the maximum central axis dose) in the inner beam region, 2 mm distance-to-agreement in the profile penumbral region, and 2% (of the maximum central axis dose) in the outer beam region.[32] Relative profile doses for the 2×2, 3×3, and 10×10 cm^2, 15 MV photon fields are presented in Figs. 6(a), 6(b), and 6(c), respectively. Profiles at each field size were acquired at depths of 3.0 cm (d_{max}), 10 cm and 20 cm. As with the 6 MV profiles, all curves have been normalized to the maximum central axis point dose. The Monte Carlo uncertainties associated with these curves are similar to those found for the 6 MV profile calculations. The differences versus measurements for the 15 MV, 2×2 cm^2 [Fig. 6(a)] and 3×3 cm^2 [Fig. 6(b)] DPM profiles are comparable to those for the corresponding 6 MV profiles, with the exception of the 10×10 cm^2 [Fig. 6(c)], where the 15 MV calculated profile penumbrae show a greater difference from measurements relative to the 6 MV beam [Fig. 5(c)]. This discrepancy is attributed to a slight misalignment in the jaw positions for the 15 MV beam. As reported by Bagheri *et al.*,[33] an uncertainty of 0.05 cm in the lateral jaw position (located at $z = 39$ cm in their study) can cause changes of up to 8% of the maximum dose in the penumbral region. Given that the measurements and calculations in this work were conducted with an estimated uncertainty of 0.1 cm (2σ) in the jaw positions, the penumbral differences are found to be within the experimental uncertainty.

B. Inhomogeneous phantom benchmarks

Figures 7(a) and 7(b) illustrate the central axis depth dose comparisons for 6 and 15 MV photons, respectively, within the inhomogeneous (composite solid water/lung/solid water) phantom. Depth doses for the 2×2, 3×3, and 10×10 cm^2 (specified at the isocenter) have all been plotted on the same graph. DPM calculations are shown in the solid lines with ion chamber measurements depicted in open markers. All curves have been normalized to the doses for the respective energies and field sizes at a depth of 10 cm in the homogeneous (water only) situation. The 2×2 and 3×3 cm^2 field sizes include scaling factors of 0.5 and 0.75, respectively, so that all curves can be viewed on the same plot. The rms deviations of the differences between calculations and measurements as well as the Monte Carlo uncertainties, at the 1σ level, are presented in Table I for the depth dose curves in Figs. 7(a) and 7(b). The rms deviations for both energies and all field sizes are within 1%, and fall within the estimated Monte Carlo uncertainty ranges. Maximum point differences, from -1% to -2.7%, are evident, however, these are deemed to be within the 1 mm experimental uncertainty with

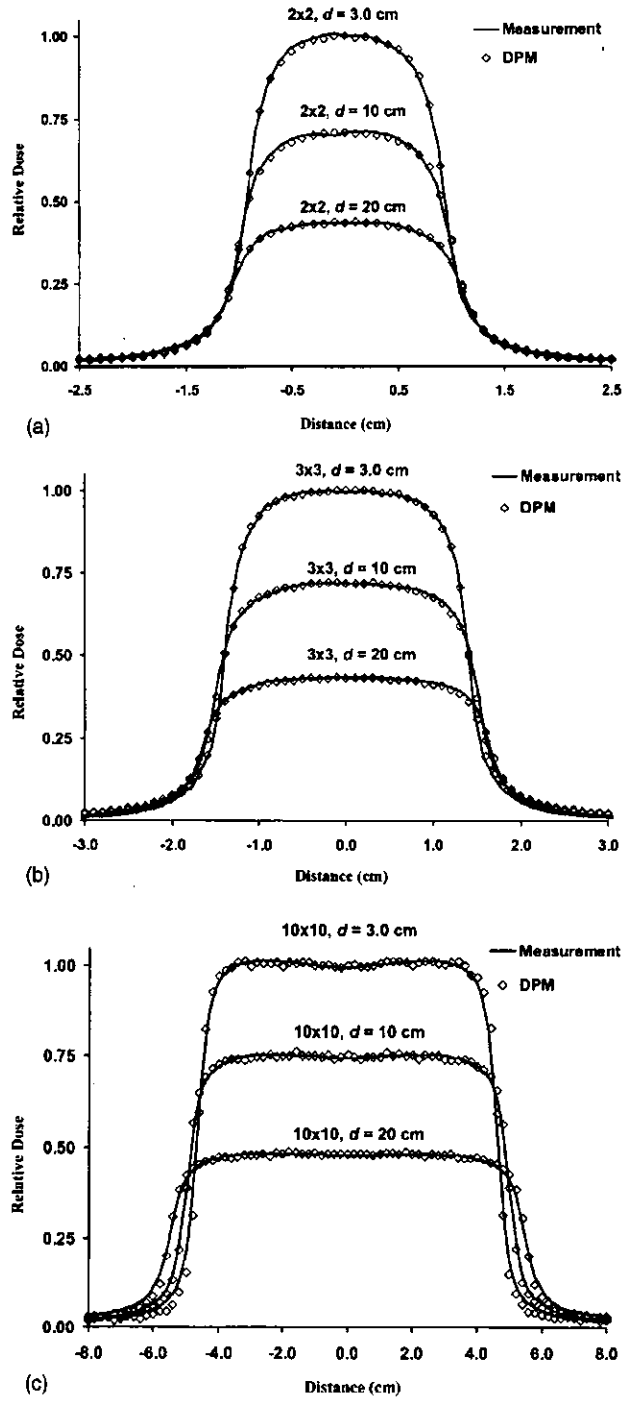


FIG. 6. Relative central axis profile doses for 15 MV photons in a water phantom. DPM calculations are shown with markers and diode measurements are shown in the solid lines. Profiles are illustrated for (a) 2×2 cm^2, (b) 3×3 cm^2, and (c) 10×10 cm^2 field sizes at depths of d_{max} (3.0 cm), 10, and 20 cm. All curves have been normalized to the maximum central axis point dose.

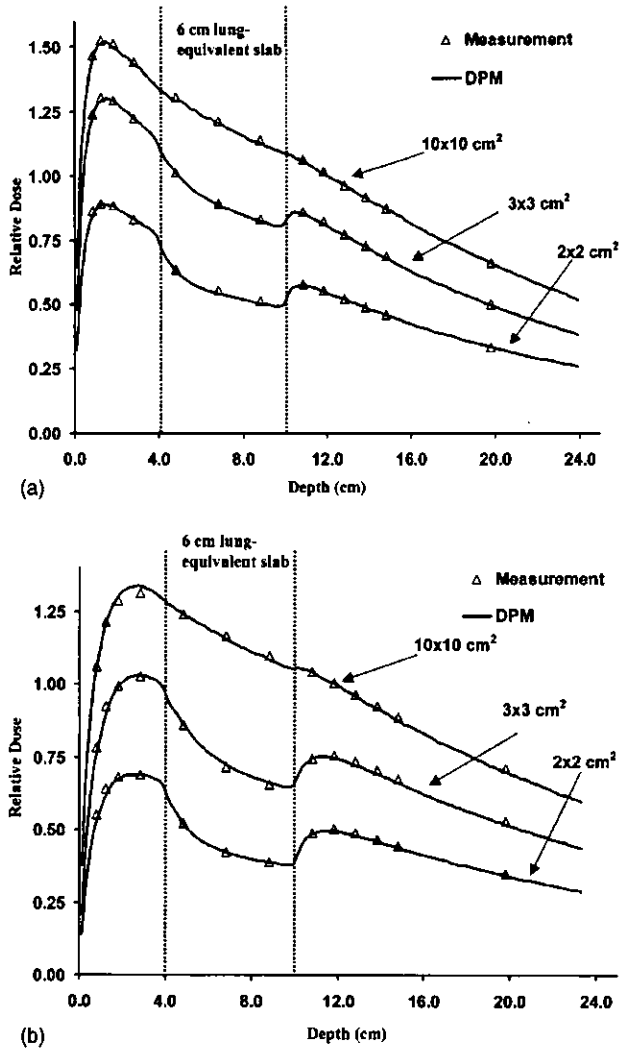


FIG. 7. Relative central axis depth dose for (a) 6 MV and (b) 15 MV photons in the inhomogeneous (solid-water/lung/solid-water) phantom. DPM calculations are shown in the solid lines and ion chamber measurements are shown with markers. Curves are illustrated for field sizes of 2×2, 3×3, and 10×10 cm^2. Depth dose curves have been normalized to the doses, for the respective field sizes, at 10 cm depth in the homogeneous phantom. The 2×2 and 3×3 cm^2 curves have been scaled using factors of 0.5 and 0.75, respectively, for an illustration on the same graph.

respect to depth positioning of the ion chamber. The AAPM Task Group No. 53[32] suggested that acceptability criteria for slab inhomogeneities along the central axis is 3%; this excludes regions of electronic disequilibrium and is therefore only applicable in our study for the largest field size (10×10 cm^2). It is clear that more suitable difference criteria are necessary for experiments involving regions of electronic disequilibrium.

Several important physical phenomena are observable in the depth dose curves within the inhomogeneous phantom presented in Fig. 7. We note a severe reduction in dose within the lung-equivalent slab, which tends to worsen as the field size is reduced. This is due to the loss of lateral electronic equilibrium at the smaller field sizes (2×2 and 3×3 cm^2), coupled with a reduction of photon scatter in the low-density medium. The loss of lateral electronic equilibrium is greater for higher energy photons because of the increased range of the lateral electrons; this is exemplified for the 15 MV photon beam in Fig. 7(b) where the reduction

in dose within the inhomogeneity at smaller field sizes is greater than that of the 6 MV beam [Fig. 7(a)]. As the field size is increased the reduction in dose within the lung becomes less pronounced, and for a 10×10 cm^2 field, the effect of the inhomogeneity is much less discernible. At depths immediately beyond the distal end of the inhomogeneity, the dose increases because of the reduced attenuation in the lung slab as well as the increase in electron backscatter from the higher-density water. Similar results have been reported by other investigators,[34,35] albeit for larger field sizes. It is instructive to present the depth dose curves in Fig. 7 in the form of lung dose correction factors (CF), where the CF is defined at a given depth on the central axis as the dose in the heterogeneous phantom divided by the dose in a solid water phantom. The CFs provide a more quantitative means of evaluating the influence of the inhomogeneity on the depth dose. CFs are illustrated as a function of depth and field size in Figs. 8(a) and 8(b) for 6 and 15 MV photons, respectively. DPM calculated CFs are shown as dashed lines with open markers, while those for the measurements are shown with solid markers. Differences between calculated and measured CFs are similar to those observed for the depth dose comparisons in Fig. 7. As seen with the depth dose curves in Fig. 7, there is a reduction in CFs (in Fig. 8) as the field size is reduced. The dose reduction is greater at 15 MV; for example, the lowest CF for a 2×2 cm^2 field size is approximately 0.85 for 6 MV [Fig. 8(a)], while it is 0.70 for 15 MV [Fig. 8(b)]; this follows from the fact that there is increased lateral electronic disequilibrium at 15 MV, as discussed earlier. Beyond the distal end of the interface the CFs increase above 1.0, indicating that the dose is higher in the solid water beyond the interface relative to that in the homogeneous situation. In addition, the CFs are higher at the smaller field sizes. These effects are attributed to the following factors: (a) reduced attenuation within the lung slab resulting in an increase in photon fluence in solid water beyond the lung relative to the homogeneous depth dose, and (b) more dose being transported away from the lung into the adjacent distal solid water at smaller field sizes, due to increased lateral electron transport issues. The correction factors in solid water beyond the lung are also found to be higher at 6 MV relative to 15 MV. This is due to the reduction in photon attenuation which results when water is replaced with lung in the composite phantom. As the average attenuation coefficient is higher at 6 MV than at 15 MV, this effect (and hence the CFs) will be larger at 6 MV in the water medium beyond the lung-equivalent slab. The CFs reported in this study are in good agreement with those measured by Rice *et al.*[34] for 4 and 15 MV photon beams.

Figures 9(a) and 9(b) show the relative profile dose comparisons in the inhomogeneous phantom for 6 and 15 MV photons, respectively. Each figure illustrates profiles for the 2×2 and 10×10 cm^2 field sizes at depths of 8 cm (within the lung slab) and 12 cm (in the solid water beyond the lung). All profiles have been normalized to the respective central axis doses; the profiles at a depth of 12 cm contain an additional scaling factor of 0.75 for representation on the same graph. DPM calculations are depicted with open markers, ion chamber measurements with closed markers and EDR film measurements with solid lines. The 1σ Monte Carlo uncertainties were, on average, less than 1% for the 2×2 cm^2 field and less than 2% for 10×10 cm^2 field size. The differences between DPM calculations and measurements are generally within (a) 2% in the inner beam region (dose >90%), (b) 1 mm distance-to-agreement in the penumbral region (20%<dose<80%) when compared with EDR film measurements, and (c) 2% in the outer beam region (dose <20%). The lateral broadening of the profile penumbral regions within the lung slab is also illustrated in Fig. 9. The profile penumbral regions are consistently broader at a depth of 8 cm due to the lateral spreading of dose within the lung; this effect is more significant for the 2×2 cm^2 profiles. An estimate of the 80%–20% penumbral widths for the 2×2 cm^2 field at depths of 8 cm and 12 cm—including an inverse-square correction so that both profiles are at the same distance from the source—gives values of 4.6 mm ($d=8$ cm) and 3 mm ($d=12$ cm), for the 6 MV beam. The corresponding penumbra values for the 15 MV beam (2

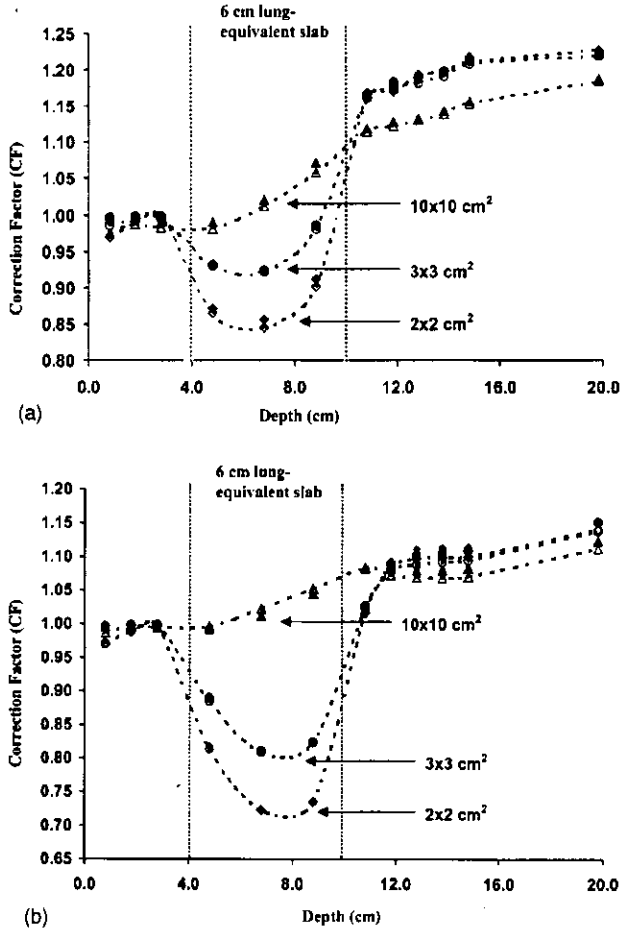


FIG. 8. Correction factors (CF) as a function of depth for the depth dose curves illustrated in Fig. 7, for (a) 6 MV and (b) 15 MV photons. The CF is defined as the ratio of dose in the inhomogeneous phantom to that in the homogeneous (water) phantom, at a given field size and depth. DPM calculated CFs are shown as solid lines with open markers and measured CFs are shown with closed markers. CFs are illustrated for field sizes of 2×2, 3×3, and 10×10 cm^2.

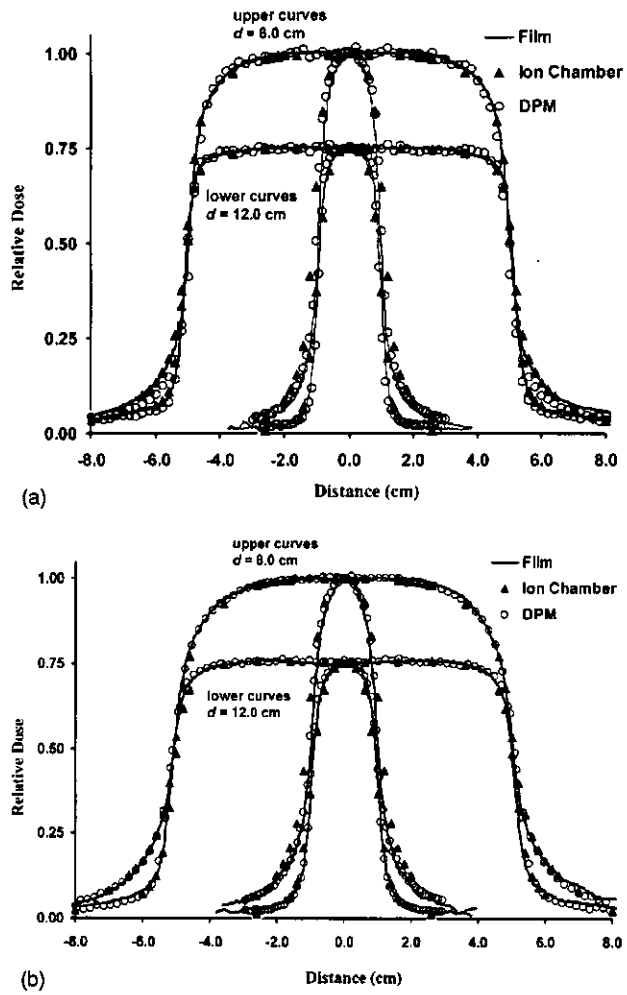


FIG. 9. Relative central axis profile doses for (a) 6 MV and (b) 15 MV photons in the inhomogeneous (solid-water/lung/solid-water) phantom. DPM calculations are shown with open markers, ion chamber measurements with closed markers, and film measurements with solid lines. Profiles are illustrated for field sizes of 2×2 and $10\times10\ \text{cm}^2$ at depths of 8 cm (within the lung) and 12 cm (within the solid water). Curves have been normalized to their respective central axis doses; profiles at a depth of 12 cm include a scaling factor of 0.75 for illustration purposes.

$\times2\ \text{cm}^2$ field) are 6.0 mm ($d=8$ cm) and 4.2 mm ($d=12$ cm). Finally, it is observed that the profile penumbral broadening for the $2\times2\ \text{cm}^2$ field within the lung is largest with the ion chamber measurements. This is due to the volume averaging effect of the ion chamber, which appears to be most prominent on the small field profiles in the lung, where the lateral scattering of electrons is significant. Given that the ion chamber has a 3 mm inner radius, it is not unexpected that the penumbral blurring (due to volume averaging) is most pronounced for the $2\times2\ \text{cm}^2$ profiles.

VI. CONCLUSION

Our goal in this investigation was to benchmark the accuracy of the DPM Monte Carlo code for photon dose calculations in a low-density, lung-equivalent medium, using small field sizes, where lateral electronic disequilibrium effects are maximized. In order to properly evaluate the radiation transport accuracy of the DPM code under the aforementioned conditions, it was necessary to first benchmark the accuracy of the phase-space calculations of the linear accelerator treatment (conducted with BEAMnrc) in a homogeneous water phantom for field sizes of 2×2, 3×3, and $10\times10\ \text{cm}^2$ (specified at the isocenter). DPM central axis depth dose and profile calculations for 2×2, 3×3, and $10\times10\ \text{cm}^2$ field sizes in water were, on average, within $\pm1\%/1$ mm relative agreement with diode measurements. The homogeneous phantom study illustrated that the phase space simulations provided an accurate description of the radiation source, the intent being that any differences noted in the inhomogeneous situation would not be due to issues regarding the radiation source. The inhomogeneous phantom experiment, conducted in a solid-water/lung/solid-water phantom, involved the use of an ion chamber and film for depth dose and profile measurements at depths within and beyond the lung. DPM calculations were generally within $\pm2\%$ relative agreement with measurements for all depth dose and profile comparisons (in the inner and outer beam regions) and within 1–2 mm distance-to-agreement in the profile penumbral regions, for all field sizes in the inhomogeneous phantom.

Although this work has demonstrated that the DPM Monte Carlo code is capable, even under conditions of severe electronic disequilibrium, of accurate photon beam dose calculations in lung-equivalent media, we realize that much more testing and validation is necessary before DPM is clinically useful. For example, the issue of accurate dose calculations and measurements in the depth dose buildup region is important and requires further investigation. This is an issue that is quite clinically relevant, especially in anatomical regions, such as the head and neck, where lesions can be superficially seated. Another important topic involves the phase space simulation of the accelerator treatment head. Sheikh-Bagheri and Rogers[36] have recently shown that accelerator head modeling is sensitive to issues such as the incident electron-on-target energy spectrum and angular distribution. Although we have found the use of mono-energetic, mono-directional electrons-on-target to be an accurate representation of the fields sizes studied in this work, it is possible that this approximation breaks down for larger field sizes (greater than $10\times10\ \text{cm}^2$). These issues along with others will be the focus of future work.

ACKNOWLEDGMENTS

This work has been supported in part by NIH Grant No. P01-CA59827 and by a University of Michigan Cancer Center grant funded by the John and Suzanne Munn Endowment. We are indebted to Dave Rogers, and the BEAM group at the NRCC, Blake Walters, in particular, for his assistance with all our questions. We are grateful to Dr. Bose of the UM Center for Advanced Computing, for his assistance with parallelizing DPM on the shared linux cluster. We would like to thank Gary Engeseth of Gammex, RMI (Gammex, RMI, Middleton, WI) for providing the atomic composition of the lung-equivalent material, and Varian Medical Systems

(Varian Associates, Palo Alto, CA) for providing proprietary information on the design of the linear accelerator.

[a]Electronic mail: indrin@med.umich.edu

[1]J. Van Dyke, R. B. Barnett, J. E. Cygler, and P. C. Shragge, "Commissioning and quality assurance of treatment planning computers," Int. J. Radiat. Oncol., Biol., Phys. **26**, 261–273 (1992).

[2]P. Andreo, "Monte Carlo techniques in medical radiation physics," Phys. Med. Biol. **26**, 861–920 (1991).

[3]T. R. Mackie, J. W. Scrimger, and J. J. Battista, "A convolution method of calculating dose for 15 MV X-rays," Med. Phys. **12**, 188–196 (1985).

[4]R. Mohan, "Why Monte Carlo?," *in Proceedings of the XIIth International Conference on the Use of Computers in Radiation Therapy*, Salt Lake City, UT, edited by D. D. Leavitt and G. Starkschall (Medical Physics Publishing, Madison, WI, 1997), pp. 16–18.

[5]M. R. Arnfield, C. Hartmann Siantar, J. Siebers, P. Garmon, L. Cox, and R. Mohan, "The impact of electron transport on the accuracy of computed dose," Med. Phys. **27**, 1266–1274 (2000).

[6]A. E. Schach von Wittenau, L. J. Cox, P. M. Bergstrom, W. P. Chandler, and C. L. Hartmann-Siantar, "Correlated histogram representation of Monte Carlo derived medical accelerator photon output phase space," Med. Phys. **26**, 1196–1212 (2000).

[7]C. L. Hartmann Siantar *et al.*, "Description and dosimetric verification of the PEREGRINE Monte Carlo dose calculation system for photon beams incident on a water phantom," Med. Phys. **28**, 1322–1337 (2001).

[8]J. V. Siebers, P. J. Keall, J. O. Kim, and R. Mohan, "Performance benchmarks of the MCV Monte Carlo system," in *Proceedings of the XIIIth International Conference on the Use of Computers in Radiation Therapy*, edited by W. Schlegel and T. Bortfeld (Springer-Verlag, Berlin, Germany, 2000), pp. 129–131.

[9]L. Wang, C. S. Chui, and M. Lovelock, "A patient-specific Monte Carlo dose-calculation method for photon beams," Med. Phys. **25**, 867–878 (1998).

[10]J. Sempau, S. J. Wilderman, and A. F. Bielajew, "DPM, a fast, accurate Monte Carlo code optimized for photon and electron radiotherapy treatment planning dose calculations," Phys. Med. Biol. **45**, 2263–2291 (2000).

[11]I. J. Chetty, J. M. Moran, T. S. Nurushev, D. L. McShan, B. A. Fraass, S. J. Wilderman, and A. F. Bielajew, "Experimental validation of the DPM Monte Carlo code using minimally scattered electron beams in heterogeneous media," Phys. Med. Biol. **47**, 1837–1851 (2002).

[12]I. Kawrakow, M. Fippel, and K. Friedrich, "3D electron dose calculation using a Voxel based Monte Carlo algorithm (VMC)," Med. Phys. **23**, 445–457 (1996).

[13]M. Fippel, "Fast Monte Carlo dose calculation for photon beams based on the VMC electron algorithm," Med. Phys. **26**, 1466–1475 (1999).

[14]J. Deng, S. B. Jiang, J. Li, T. Pawlicki, and C. M. Ma, "Photon beam characterization and modeling for Monte Carlo treatment planning," Phys. Med. Biol. **45**, 411–427 (2000).

[15]C. M. Ma *et al.*, "Clinical implementation of a Monte Carlo treatment planning system," Med. Phys. **26**, 2133–2143 (1999).

[16]C. M. Ma, P. Reckwerdt, M. Holmes, D. W. O. Rogers, and B. Geiser, "DOSXYZ Users Manual," NRCC Report PIRS 509B(rev C), 1998.

[17]J. J. DeMarco, T. D. Solberg, and J. B. Smathers, "A CT-based Monte Carlo dosimetry tool for radiotherapy treatment planning and analysis," Med. Phys. **25**, 1–11 (1998).

[18]M. K. Fix, M. Stampanoni, P. Manser, E. J. Born, R. Mini, and P. Ruegsegger, "A multiple source model for 6 MV photon beam dose calculations," Phys. Med. Biol. **46**, 1407–1428 (2001).

[19]P. J. Keall and P. W. Hoban, "Super-Monte Carlo: A 3D electron beam dose calculation algorithm," Med. Phys. **23**, 2023–2034 (1996).

[20]H. Neuenschwander, T. R. Mackie, and P. J. Reckwerdt, "MMC—a high-performance Monte Carlo code for electron beam treatment planning," Phys. Med. Biol. **40**, 543–574 (1995).

[21]J. Sempau, A. Sánchez-Reyes, F. Salvat, H. Oulad ben Tahar, S. B. Jiang, and T. R. Mackie, "Monte Carlo simulation of electron beams from an accelerator head using PENELOPE," Phys. Med. Biol. **46**, 1163–1186 (2001).

[22]J. M. Fernández-Varea, R. Mayol, J. Baró, and F. Salvat, "On the theory and simulation of multiple elastic scattering of electrons," Nucl. Instrum. Methods Phys. Res. B **73**, 447–473 (1993).

[23]D. W. O. Rogers, D. A. Faddegon, G. X. Ding, C. M. Ma, J. We, and T. R. Mackie, "BEAM: A Monte Carlo code to simulate radiotherapy treatment units," Med. Phys. **22**, 503–524 (1995).

[24]I. Kawrakow, "Accurate condensed history Monte Carlo simulation of electron transport. I. EGSnrc, the new EGS4 version," Med. Phys. **27**, 485–498 (2000).

[25]D. M. J. Lovelock, C. S. Chui, and R. Mohan, "A Monte Carlo model of photon beams used in radiation therapy," Med. Phys. **22**, 1387–1394 (1995).

[26]I. J. Chetty and P. M. Charland, "Investigation of Kodak extended dose range (EDR) film for megavoltage photon beam dosimetry," Phys. Med. Biol. **47**, 3629–3641 (2002).

[27]J. F. Briesmeister, "MCNP™—A general Monte Carlo *N* Particle transport code," Los Alamos National Laboratory Report LA-12625-M, 1997.

[28]B. R. B. Walters, I. Kawrakow, and D. W. O. Rogers, "History by history statistical estimators in the BEAM code system," Med. Phys. **29**, 2745–2752 (2002).

[29]C. S. Reft, "Output calibration in solid water for high energy photon beams," Med. Phys. **16**, 299–301 (1988).

[30]G. X. Ding, "Dose discrepancies between Monte Carlo calculations and measurements in the buildup region for a high-energy photon beam," Med. Phys. **29**, 2459–2463 (2002).

[31]S. Yokoyama, P. Roberson, J. Moran, D. Litzenberg, and B. Fraass, "Buildup region dependence on photon dose delivery technique for IMRT," Med. Phys. **29**, 1315 (2002) (abstract).

[32]B. Fraass, K. Doppke, M. Hunt, G. Kutcher, G. Starkschall, R. Stern, and J. Van Dyke, "American Association of Physicists in Medicine Radiation Therapy Committee Task Group 53: Quality assurance for clinical radiotherapy treatment planning," Med. Phys. **25**, 1773–1829 (1998).

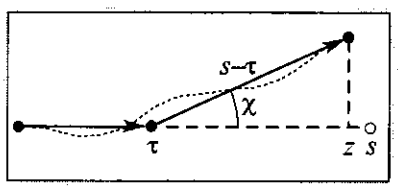
[33]D. Sheikh-Bagheri *et al.*, "Comparison of measured and Monte Carlo calculated dose distributions from the NRC linac," Med. Phys. **27**, 2256–2266 (2000).

[34]R. Rice, J. L. Hansen, L. M. Chin, B. J. Mijnheer, and B. E. Bjärngard, "The influence of ionization chamber and phantom design on the measurement of lung dose in photon beams," Med. Phys. **15**, 884–890 (1988).

[35]E. El-Khatib, M. Evans, M. Pla, and J. R. Cunningham, "Evaluation of lung dose correction methods for photon irradiations of thorax phantoms," Int. J. Radiat. Oncol., Biol., Phys. **17**, 871–878 (1989).

[36]D. Sheikh-Bagheri and D. W. O. Rogers, "Sensitivity of megavoltage photon beam Monte Carlo simulations to electron beam and other parameters," Med. Phys. **29**, 379–390 (2002).

PENELOPE – A Code System for Monte Carlo Simulation of Electron and Photon Transport



Workshop Proceedings
Issy-les-Moulineaux, France
5-7 November 2001

AEN NEA

NUCLEAR • ENERGY • AGENCY

NUCLEAR SCIENCE COMMITTEE

PENELOPE

A Code System for Monte Carlo Simulation of Electron and Photon Transport

Francesc Salvat, José M. Fernández-Varea
Facultat de Física (ECM)
Universitat de Barcelona
Spain

Eduardo Acosta
Facultad De Matemática, Astronomía y Física
Universidad Nacional de Córdoba
Argentina

Josep Sempau
Institut de Tècniques Energètiques
Universitat Politècnica de Catalunya
Spain

NUCLEAR ENERGY AGENCY
ORGANISATION FOR ECONOMIC CO-OPERATION AND DEVELOPMENT

ORGANISATION FOR ECONOMIC CO-OPERATION AND DEVELOPMENT

Pursuant to Article 1 of the Convention signed in Paris on 14th December 1960, and which came into force on 30th September 1961, the Organisation for Economic Co-operation and Development (OECD) shall promote policies designed:

- to achieve the highest sustainable economic growth and employment and a rising standard of living in Member countries, while maintaining financial stability, and thus to contribute to the development of the world economy;
- to contribute to sound economic expansion in Member as well as non-member countries in the process of economic development; and
- to contribute to the expansion of world trade on a multilateral, non-discriminatory basis in accordance with international obligations.

The original Member countries of the OECD are Austria, Belgium, Canada, Denmark, France, Germany, Greece, Iceland, Ireland, Italy, Luxembourg, the Netherlands, Norway, Portugal, Spain, Sweden, Switzerland, Turkey, the United Kingdom and the United States. The following countries became Members subsequently through accession at the dates indicated hereafter: Japan (28th April 1964), Finland (28th January 1969), Australia (7th June 1971), New Zealand (29th May 1973), Mexico (18th May 1994), the Czech Republic (21st December 1995), Hungary (7th May 1996), Poland (22nd November 1996), Korea (12th December 1996) and the Slovak Republic (14 December 2000). The Commission of the European Communities takes part in the work of the OECD (Article 13 of the OECD Convention).

NUCLEAR ENERGY AGENCY

The OECD Nuclear Energy Agency (NEA) was established on 1st February 1958 under the name of the OEEC European Nuclear Energy Agency. It received its present designation on 20th April 1972, when Japan became its first non-European full Member. NEA membership today consists of 27 OECD Member countries: Australia, Austria, Belgium, Canada, Czech Republic, Denmark, Finland, France, Germany, Greece, Hungary, Iceland, Ireland, Italy, Japan, Luxembourg, Mexico, the Netherlands, Norway, Portugal, Republic of Korea, Spain, Sweden, Switzerland, Turkey, the United Kingdom and the United States. The Commission of the European Communities also takes part in the work of the Agency.

The mission of the NEA is:

- to assist its Member countries in maintaining and further developing, through international co-operation, the scientific, technological and legal bases required for a safe, environmentally friendly and economical use of nuclear energy for peaceful purposes, as well as
- to provide authoritative assessments and to forge common understandings on key issues, as input to government decisions on nuclear energy policy and to broader OECD policy analyses in areas such as energy and sustainable development.

Specific areas of competence of the NEA include safety and regulation of nuclear activities, radioactive waste management, radiological protection, nuclear science, economic and technical analyses of the nuclear fuel cycle, nuclear law and liability, and public information. The NEA Data Bank provides nuclear data and computer program services for participating countries.

In these and related tasks, the NEA works in close collaboration with the International Atomic Energy Agency in Vienna, with which it has a Co-operation Agreement, as well as with other international organisations in the nuclear field.

FOREWORD

The OECD/NEA Data Bank was established in order to promote the effective sharing of data and software developed in Member countries in the field of nuclear technology and radiation physics applications. It operates a Computer Program Service (CPS) related to nuclear energy applications. The software library collects programs, compiles and verifies them in an appropriate computer environment, ensuring that the computer program package is complete and adequately documented. Internationally agreed quality assurance methods are used in the verification process.

In order to obtain good results in modelling the behaviour of technological systems two conditions must be fulfilled:

1. Good quality and validated computer codes and associated basic data libraries should be used.

2. Modelling should be done by a qualified user of such codes.

One area to which special effort has been devoted in recent years is radiation transport. Workshops and training courses including the use of computer codes have been organised in the field of neutral particle transport for codes using both deterministic and stochastic methods. The area of charged particle transport, and in particular electron-photon transport, has received increased attention for a number of technological and medical applications. At the most recent Monte Carlo 2000 Conference, held on 23-26 October 2000 in Lisbon, Portugal, about half of the papers covered electron-photon transport and its application.

A new computer code was recently released to the NEA Data Bank for general distribution: "PENELOPE, A Code System for Monte Carlo Simulation of Electron and Photon Transport" developed by Francesc Salvat, José M. Fernández-Varea, Eduardo Acosta and Josep Sempau. This code began to be widely used by radiation physicists, who requested that a workshop with hands-on training with the PENELOPE code be organised. The NEA Nuclear Science Committee endorsed this request and the authors agreed to teach a course covering the physics behind the code and to demonstrate, with corresponding exercises, how it can be used for practical applications.

These proceedings contain the teaching notes of the workshop and training course held in November 2001.

Abstract

The computer code system PENELOPE (version 2001) performs Monte Carlo simulation of coupled electron-photon transport in arbitrary materials for a wide energy range, from a few hundred eV to about 1 GeV. Photon transport is simulated by means of the standard, detailed simulation scheme. Electron and positron histories are generated on the basis of a mixed procedure, which combines detailed simulation of hard events with condensed simulation of soft interactions. A geometry package called PENGEOM permits the generation of random electron-photon showers in material systems consisting of homogeneous bodies limited by quadric surfaces, i.e. planes, spheres, cylinders, etc. This report is intended not only to serve as a manual of the PENELOPE code system, but also to provide the user with the necessary information to understand the details of the Monte Carlo algorithm.

Keywords: *Radiation transport, electron-photon showers, Monte Carlo simulation, sampling algorithms, quadric geometry*

Symbols and numerical values of constants frequently used in the text (Mohr and Taylor, 2000)

Quantity	Symbol	Value
Avogadro's number	N_A	6.022142×10^{23} mol^{-1}
Velocity of light in vacuum	c	2.997925×10^{8} m s^{-1}
Reduced Planck's constant	$\hbar = h/(2\pi)$	6.582119×10^{-22} MeV s
Electron charge	e	1.602176×10^{-19} C
Electron mass	m_e	9.109382×10^{-31} kg
Electron rest energy	$m_e c^2$	0.5109989 MeV
Classical electron radius	$r_e = e^2/(m_e c^2)$	2.817940×10^{-15} m
Fine-structure constant	$\alpha = e^2/(\hbar c)$	1/137.0360
Bohr radius	$a_0 = \hbar^2/(m_e e^2)$	$0.5291772 \times 10^{-10}$ m

TABLE OF CONTENTS

PREFACE

Radiation transport in matter has been a subject of intense work since the beginning of the 20th century. Today, we know that high-energy photons, electrons and positrons penetrating matter suffer multiple interactions by which energy is transferred to the atoms and molecules of the material and secondary particles are produced[1]. By repeated interaction with the medium, a high-energy particle originates a cascade of particles which is usually referred to as a shower. After each interaction of a particle, its energy is reduced and further particles may be generated so that the evolution of the shower represents an effective degradation in energy. As time goes on, the initial energy is progressively deposited into the medium, while that remaining is shared by an increasingly larger number of particles.

A reliable description of shower evolution is required in a number of fields. Thus, knowledge of radiation transport properties is needed for quantitative analysis in surface electron spectroscopies (Jablonski, 1987; Tofterup, 1986), positron surface spectroscopy (Shultz and Lynn, 1988), electron microscopy (Reimer, 1985), electron energy loss spectroscopy (Reimer, *et al.*, 1992), electron probe microanalysis (Heinrich and Newbury, 1991), etc. Detailed information on shower evolution is also required for the design and quantitative use of radiation detectors (Titus, 1970; Berger and Seltzer, 1972). A field where radiation transport studies play an important sociological role is that of radiation dosimetry and radiotherapy (Andreo, 1991).

The study of radiation transport problems was initially attempted on the basis of the Boltzmann transport equation. However, this procedure comes up against considerable difficulties when applied to limited geometries, with the result that numerical methods based on the transport equation have only had a certain success in simple geometries, mainly for unlimited and semi-infinite media (see e.g. Zheng-Ming and Brahme, 1993). At the end of the 1950s, with the availability of computers, Monte Carlo simulation methods were developed as a powerful alternative to deal with transport problems. Basically, the evolution of an electron-photon shower is of a random nature, so that this is a process particularly amenable to Monte Carlo simulation. Detailed simulation, where all the interactions experienced by a particle are simulated in chronological succession, is exact, i.e. it yields the same results as the rigorous solution of the transport equation (apart from the inherent statistical uncertainties).

To our knowledge, the first numerical Monte Carlo simulation of photon transport is that of Hayward and Hubbell (1954) who generated 67 photon histories using a desk calculator. The simulation of photon transport is straightforward since the mean number of events in each history is fairly small. Indeed, the photon is effectively absorbed after a single photoelectric or pair-production interaction or after a few Compton interactions (say, of the order of 10). With present-day computational facilities, detailed simulation of photon transport is a simple routine task.

The simulation of electron and positron transport is much more difficult than that of photons. The main reason is that the average energy loss of an electron in a single interaction is very small (of the order of a few tens of eV). As a consequence, high-energy electrons suffer a large number of interactions before being effectively absorbed in the medium. In practice, detailed simulation is feasible

[1] In this report, the term particle will be used to designate either photons, electrons or positrons.

only when the average number of collisions per track is not too large (say, up to a few hundred). Experimental situations which are amenable to detailed simulation are those involving either electron sources with low initial kinetic energies (up to about 100 keV) or special geometries such as electron beams impinging on thin foils. For larger initial energies, and thick geometries, the average number of collisions experienced by an electron until it is effectively stopped becomes very large, and detailed simulation is very inefficient.

For high-energy electrons and positrons, most of the Monte Carlo codes currently available [e.g. ETRAN (Berger and Seltzer, 1988), ITS3 (Halbleib, *et al.*, 1992), EGS4 (Nelson, *et al.*, 1985), EGSnrc (Kawrakow and Rogers, 2000), GEANT (Brun, *et al.*, 1986), MCNP (Briesmeister, 1997), ...] have recourse to multiple scattering theories which allow the simulation of the global effect of a large number of events in a track segment of a given length (step). Following Berger (1963), these simulation procedures will be referred to as "condensed" Monte Carlo methods. The multiple scattering theories implemented in condensed simulation algorithms are only approximate and may lead to systematic errors, which can be made evident by the dependence of the simulation results on the adopted step length (Bielajew and Rogers, 1987). To analyse their magnitude, one can perform simulations of the same arrangement with different step lengths. The results are usually found to stabilise when the step length is reduced, while computation time increases rapidly, roughly in proportion to the inverse of the step length. Thus, for each particular problem, one must reach a certain compromise between available computer time and attainable accuracy. It is also worth noting that, owing to the nature of certain multiple scattering theories and/or to the particular way they are implemented in the simulation code, the use of very short step lengths may introduce spurious effects in the simulation results. For instance, the multiple elastic scattering theory of Molière (1948) which is the model used in EGS4-based codes, is not applicable to step lengths shorter than a few times the elastic mean free path (see e.g. Fernández-Varea, *et al.*, 1993b) and multiple elastic scattering has to be switched off when the step length becomes smaller than this value. As a consequence, stabilisation for short step lengths does not necessarily imply that simulation results are correct. Condensed schemes also have difficulties in generating particle tracks in the vicinity of an interface, i.e. a surface separating two media of different compositions. When the particle moves near an interface, the step length must be kept smaller than the minimum distance to the interface so as to make sure that the step is completely contained in the initial medium (Bielajew and Rogers, 1987). This may complicate the code considerably, even for relatively simple geometries.

In the present report, we describe the version 2001 of PENELOPE, a Monte Carlo algorithm and computer code for the simulation of coupled electron-photon transport. The name is an acronym that stands for PENetration and Energy LOss of Positrons and Electrons (photon simulation was introduced later). The simulation algorithm is based on a scattering model that combines numerical databases with analytical cross section models for the different interaction mechanisms and is applicable to energies (kinetic energies in the case of electrons and positrons) from a few hundred eV to ~1 GeV. Photon transport is simulated by means of the conventional detailed method. The simulation of electron and positron transport is performed by means of a mixed procedure. Hard interactions, with scattering angle θ and/or energy loss W greater than pre-selected cut-off values θ_c and W_c, are simulated in detail. Soft interactions, with scattering angle or energy loss less than the corresponding cut-offs, are described by means of multiple scattering approaches. This simulation scheme handles lateral displacements and interface crossing appropriately and provides a consistent description of energy straggling. The simulation is stable under variations of the cut-offs θ_c, W_c and these can be made quite large, thus speeding up the calculation considerably, without altering the results. A characteristic feature of our code is that the most delicate parts of the simulation are handled internally; electrons, positrons and photons are simulated by calling the same subroutines. Thus, from the user's point of view, PENELOPE makes the practical simulation of electrons and positrons as simple as that of photons (although simulating a charged particle may take a longer time).

The present version of PENELOPE is the result of continued evolution from the first version, which was released in 1996. The idea of developing a general-purpose Monte Carlo code, with better modelling than those available at that time, arose during a short course on radiation transport simulation given by F. Salvat at the Radiation Metrology Unit, CIEMAT (Madrid), in 1988. The present version 2001 contains substantial changes/improvements to the previous versions 1996 and 2000. As for the physics, the model for electron/positron elastic scattering has been revised, bremsstrahlung emission is now simulated using partial-wave data instead of analytical approximate formulae, photoelectric absorption in K and L-shells is described from the corresponding partial cross sections, and fluorescence radiation from vacancies in K and L-shells is now followed. Refinements have also been introduced in the electron/positron transport mechanics, mostly to account for the energy dependence of the mean free paths for hard events. The simulation routines have been re-programmed in a more structured (and readable) way and new example MAIN programs have been written, with a more flexible input and expanded output.

This report is intended not only to serve as a manual of the simulation package, but also to provide the user with the necessary information to understand the details of the Monte Carlo algorithm. In Chapter 1 we give a brief survey of random sampling methods and an elementary introduction to Monte Carlo simulation of radiation transport. The cross sections adopted in PENELOPE to describe particle interactions, and the associated sampling techniques, are presented in Chapters 2 and 3. Chapter 4 is devoted to mixed simulation methods for electron and positron transport. In Chapter 5, a relatively simple, but effective, method to handle simulation in quadric geometries is presented. The FORTRAN77 simulation package PENELOPE and other complementary programs, are described in Chapter 6, which also provides instructions to operate them. Information on relativistic kinematics and numerical methods is given in Appendices A and B. Finally, Appendix C is devoted to simulation of electron/positron transport under external, static electric and magnetic fields. The source files of PENELOPE, the auxiliary programs and the database are supplied on a ZIP-compressed file, which is distributed by the NEA Data Bank[2] and the RSICC[3]. The code is also available from the authors, but we would appreciate it if users did try to get the code from these institutions.

In the course of our Monte Carlo research, we have had the fortune of getting much help from numerous friends and colleagues. Since the mid 1980s, we have benefited from discussions with D. Liljequist, which gave shape to our first algorithm for simulation of electrons and positrons. We are particularly grateful to A. Riveros for his enthusiastic and friendly support over the years, and for guiding us into the field of microanalysis and X-ray simulation. A. Sánchez-Reyes and E. García-Toraño were the first external users of the code system; they suffered the inconveniences of using continuously changing preliminary versions of the code without complaining too much. More recently, stimulating collaboration with A.F. Bielajew has led to substantial improvements in the electron transport mechanics and in the code organisation. We are deeply indebted to J.H. Hubbell and D.E. Cullen for kindly providing us with updated information on photon interaction and atomic relaxation data. Thanks are also due to S.M. Seltzer for sending us his bremsstrahlung energy-loss database. L. Sorbier generously prepared most of the photoelectric and atomic relaxation database files and worked on the associated sampling algorithms. We are indebted to many colleagues, especially P. Andreo, for comments and suggestions, which have been of much help to improve the present version of the code. Our most sincere appreciation to the members of our research group; X. Llovet, M. Dingfelder,

[2] OECD Nuclear Energy Agency Data Bank. Le Seine Saint-Germain, 12 Boulevard des Iles. 92130 Issy-les-Moulineaux, France. E-mail: nea@nea.fr; http://www.nea.fr

[3] Radiation Safety Information Computational Center. P.O. Box 2008, Oak Ridge, TN 37831-6362, USA. E-mail: pdc@ornl.gov; http://www-rsicc.ornl.gov

J. Asenjo and C. Campos. They did much more than chasing bugs through the programs and in this write-up. Finally, we would like to thank the staff of the NEA Data Bank, particularly E. Sartori, for kindly organising the first training course on PENELOPE.

Partial support from the Fondo de Investigación Sanitaria (Ministerio de Sanidad y Consumo, Spain), projects no. 00/1070 and 01/0093, is gratefully acknowledged.

Barcelona, November 2001

Chapter 1

Monte Carlo simulation. Basic concepts

The name "Monte Carlo" was coined in the 1940s by scientists working on the nuclear weapon project in Los Alamos to designate a class of numerical methods based on the use of random numbers. Nowadays, Monte Carlo methods are widely used to solve complex physical and mathematical problems (James, 1980; Rubinstein, 1981; Kalos and Whitlock, 1986), particularly those involving multiple independent variables where more conventional numerical methods would demand formidable amounts of memory and computer time. The book by Kalos and Whitlock (1986) gives a readable survey of Monte Carlo techniques, including simple applications in radiation transport, statistical physics, and many-body quantum theory.

In Monte Carlo simulation of radiation transport, the history (track) of a particle is viewed as a random sequence of free flights that end with an interaction event where the particle changes its direction of movement, loses energy and, occasionally, produces secondary particles. The Monte Carlo simulation of a given experimental arrangement (e.g. an electron beam, coming from an accelerator and impinging on a water phantom) consists of the numerical generation of random histories. To simulate these histories we need an "interaction model", i.e. a set of differential cross sections (DCS) for the relevant interaction mechanisms. The DCSs determine the probability distribution functions (PDF) of the random variables that characterize a track; 1) free path between successive interaction events, 2) kind of interaction taking place and 3) energy loss and angular deflection in a particular event (and initial state of emitted secondary particles, if any). Once these PDFs are known, random histories can be generated by using appropriate sampling methods. If the number of generated histories is large enough, quantitative information on the transport process may be obtained by simply averaging over the simulated histories.

The Monte Carlo method yields the same information as the solution of the Boltzmann transport equation, with the same interaction model, but is easier to implement (Berger, 1963). In particular, the simulation of radiation transport in finite samples is

straightforward, while even the simplest finite geometries (e.g. thin foils) are very difficult to be dealt with by the transport equation. The main drawback of the Monte Carlo method lies in its random nature, all the results are affected by statistical uncertainties, which can be reduced at the expense of increasing the sampled population and, hence, the computation time. Under special circumstances, the statistical uncertainties may be lowered by using variance-reduction techniques (Rubinstein, 1981; Bielajew and Rogers, 1988).

1.1 Elements of probability theory

The essential characteristic of Monte Carlo simulation is the use of random numbers and random variables. A random variable is a quantity that results from a repeatable process and whose actual values (realizations) cannot be predicted with certainty. In the real world, randomness originates either from uncontrolled factors (as occurs e.g. in games of chance) or from the quantum nature of microscopic systems and processes (e.g. nuclear disintegration and radiation interactions). As a familiar example, assume that we throw two dice in a box; the sum of points in their upper faces is a discrete random variable, which can take the values 2 to 12, while the distance x between the dice is a continuous random variable, which varies between zero (dice in contact) and a maximum value determined by the dimensions of the box. In the computer, random variables are generated by means of numerical transformations of random numbers (see below).

Let x be a continuous random variable that takes values in the interval $x_{\min} \leq x \leq x_{\max}$. To measure the likelihood of obtaining x in an interval (a,b) we use the probability $\mathrm{P}\{x|a < x < b\}$, defined as the ratio n/N of the number n of values of x that fall within that interval and the total number N of generated x-values, in the limit $N \to \infty$. The probability of obtaining x in a differential interval of length $\mathrm{d}x$ about x_1 can be expressed as

$$\mathrm{P}\{x|x_1 < x < x_1 + \mathrm{d}x\} = p(x_1)\,\mathrm{d}x, \tag{1.1}$$

where $p(x)$ is the probability distribution function (PDF) of x. Since 1) negative probabilities have no meaning and 2) the obtained value of x must be somewhere in $(x_{\min},x_{\max})$, the PDF must be definite positive and normalized to unity

$$p(x) \geq 0 \qquad \text{and} \qquad \int_{x_{\min}}^{x_{\max}} p(x)\,\mathrm{d}x = 1. \tag{1.2}$$

Any "function" that satisfies these two conditions can be interpreted as a PDF. In Monte Carlo simulation we shall frequently use the uniform distribution,

$$U_{x_{\min},x_{\max}}(x) \equiv \begin{cases} 1/(x_{\max} - x_{\min}) & \text{if } x_{\min} < x < x_{\max}, \\ 0 & \text{otherwise,} \end{cases} \tag{1.3}$$

which is discontinuous. The definition (1.2) also includes singular distributions such as the Dirac delta, $\delta(x - x_0)$, which is defined by the property

$$\int_a^b f(x)\delta(x - x_0)\,\mathrm{d}x = \begin{cases} f(x_0) & \text{if } a < x_0 < b, \\ 0 & \text{if } x_0 < a \text{ or } x_0 > b \end{cases} \tag{1.4}$$

for any function $f(x)$ that is continuous at x_0. An equivalent, more intuitive definition is the following,

$$\delta(x - x_0) \equiv \lim_{\Delta \to 0} U_{x_0 - \Delta, x_0 + \Delta}(x), \tag{1.4'}$$

which represents the delta distribution as the zero-width limit of a sequence of uniform distributions centred at the point x_0. Hence, the Dirac distribution describes a single-valued discrete random variable (i.e. a constant). The PDF of a random variable x that takes the discrete values $x = x_1, x_2, \ldots$ with point probabilities $p_1, p_2, \ldots$ can be expressed as a mixture of delta distributions,

$$p(x) = \sum_i p_i\, \delta(x - x_i). \tag{1.5}$$

Discrete distributions can thus be regarded as particular forms of continuous distributions.

Given a continuous random variable x, the cumulative distribution function of x is defined by

$$\mathcal{P}(x) \equiv \int_{x_{\min}}^{x} p(x')\,\mathrm{d}x'. \tag{1.6}$$

This is a non-decreasing function of x that varies from $\mathcal{P}(x_{\min}) = 0$ to $\mathcal{P}(x_{\max}) = 1$. In the case of a discrete PDF of the form (1.5), $\mathcal{P}(x)$ is a step function. Notice that the probability $\mathrm{P}\{x | a < x < b\}$ of having x in the interval (a,b) is

$$\mathrm{P}\{x |\, a < x < b\} = \int_a^b p(x)\,\mathrm{d}x = \mathcal{P}(b) - \mathcal{P}(a), \tag{1.7}$$

and that $p(x) = \mathrm{d}\mathcal{P}(x)/\mathrm{d}x$.

The n-th moment of $p(x)$ is defined as

$$\langle x^n \rangle = \int_{x_{\min}}^{x_{\max}} x^n\, p(x)\,\mathrm{d}x. \tag{1.8}$$

The moment $\langle x^0 \rangle$ is simply the integral of $p(x)$, which is equal to unity, by definition. However, higher order moments may or may not exist. An example of a PDF that has no even-order moments is the Lorentz or Cauchy distribution,

$$p_{\mathrm{L}}(x) \equiv \frac{1}{\pi} \frac{\gamma}{\gamma^2 + x^2}, \qquad -\infty < x < \infty. \tag{1.9}$$

Its first moment, and other odd-order moments, can be assigned a finite value if they are defined as the "principal value" of the integrals, e.g.

$$\langle x \rangle_{\mathrm{L}} = \lim_{a \to \infty} \int_{-a}^{+a} x \frac{1}{\pi} \frac{\gamma}{\gamma^2 + x^2}\,\mathrm{d}x = 0, \tag{1.10}$$

but the second and higher even-order moments are infinite, irrespective of the way they are defined.

The first moment, when it exists, is called the mean or expected value of the random variable x,

$$\langle x \rangle = \int x\, p(x)\, \mathrm{d}x. \tag{1.11}$$

The expected value of a function $f(x)$ is defined in a similar way,

$$\langle f(x) \rangle \equiv \int f(x)\, p(x)\, \mathrm{d}x. \tag{1.12}$$

Since $f(x)$ is a random variable, it has its own PDF, $\pi(f)$, which is such that the probability of having f in a certain interval of length $\mathrm{d}f$ is equal to the probability of having x in the corresponding interval or intervals[1]. Thus, if $f(x)$ is a monotonously increasing function of x (so that there is a one-to-one correspondence between the values of x and f), $p(x)\,\mathrm{d}x = \pi(f)\,\mathrm{d}f$ and

$$\pi(f) = p(x)\,(\mathrm{d}f/\mathrm{d}x)^{-1}. \tag{1.13}$$

It can be shown that the definitions (1.11) and (1.12) are equivalent. If $f(x)$ increases monotonously with x, the proof is trivial: we can start from the definition (1.11) and write

$$\langle f \rangle = \int f\, \pi(f)\, \mathrm{d}f = \int f(x)\, p(x)\, (\mathrm{d}x/\mathrm{d}f)\ \mathrm{d}f = \int f(x)\, p(x)\, \mathrm{d}x,$$

which agrees with (1.12). Notice that the expectation value is linear, i.e.

$$\langle a_1 f_1(x) + a_2 f_2(x) \rangle = a_1 \langle f_1(x) \rangle + a_2 \langle f_2(x) \rangle, \tag{1.14}$$

where a_1 and a_2 are arbitrary real constants.

If the first and second moments of the PDF $p(x)$ exist, we define the variance of x [or of $p(x)$] by

$$\mathrm{var}(x) \equiv \langle (x - \langle x \rangle)^2 \rangle = \int (x - \langle x \rangle)^2\, p(x)\, \mathrm{d}x = \langle x^2 \rangle - \langle x \rangle^2. \tag{1.15}$$

The square root of the variance, $\sigma \equiv [\mathrm{var}(x)]^{1/2}$, is called the "standard deviation" (and sometimes the "standard uncertainty"); it gives a measure of the dispersion of the random variable (i.e. of the width of the PDF). The Dirac delta is the only PDF that has zero variance. Similarly, the variance of a function $f(x)$ is defined as

$$\mathrm{var}\{f(x)\} = \langle f^2(x) \rangle - \langle f(x) \rangle^2. \tag{1.16}$$

Thus, for a constant $f(x) = a$, $\langle f \rangle = a$ and $\mathrm{var}\{f\} = 0$.

[1] When $f(x)$ does not increase or decrease monotonously with x, there may be multiple values of x corresponding to a given value of f.

1.1.1 Two-dimensional random variables

Let us now consider the case of a two-dimensional random variable, (x, y). The corresponding (joint) PDF $p(x, y)$ satisfies the conditions

$$p(x, y) \geq 0 \qquad \text{and} \qquad \int \mathrm{d}x \int \mathrm{d}y\, p(x, y) = 1. \tag{1.17}$$

The *marginal* PDFs of x and y are defined as

$$q(x) \equiv \int p(x, y)\, \mathrm{d}y \qquad \text{and} \qquad q(y) \equiv \int p(x, y)\, \mathrm{d}x, \tag{1.18}$$

i.e. $q(x)$ is the probability of obtaining the value x and *any* value of y. The joint PDF can be expressed as

$$p(x, y) = q(x)\, p(y|x) = q(y)\, p(x|y), \tag{1.19}$$

where

$$p(x|y) = \frac{p(x, y)}{q(y)} \quad \text{and} \quad p(y|x) = \frac{p(x, y)}{q(x)} \tag{1.20}$$

are the *conditional* PDFs of x and y, respectively. Notice that $p(x|y)$ is the normalized PDF of x for a fixed value of y.

The expectation value of a function $f(x, y)$ is

$$\langle f(x, y) \rangle = \int \mathrm{d}x \int \mathrm{d}y\, f(x, y)\, p(x, y). \tag{1.21}$$

The moments of the PDF are defined by

$$\langle x^n y^m \rangle = \int \mathrm{d}x \int \mathrm{d}y\, x^n y^m\, p(x, y). \tag{1.22}$$

In particular,

$$\langle x^n \rangle = \int \mathrm{d}x \int \mathrm{d}y\, x^n\, p(x, y) = \int x^n q(x)\, \mathrm{d}x. \tag{1.23}$$

Again, the only moment that is necessarily defined is $\langle x^0 y^0 \rangle = 1$. When the corresponding moments exist, the variances of x and y are given by

$$\mathrm{var}(x) = \langle x^2 \rangle - \langle x \rangle^2 \quad \text{and} \quad \mathrm{var}(y) = \langle y^2 \rangle - \langle y \rangle^2. \tag{1.24}$$

The variance of $x + y$ is

$$\mathrm{var}(x + y) = \langle (x + y)^2 \rangle - \langle x + y \rangle^2 = \mathrm{var}(x) + \mathrm{var}(y) + 2\,\mathrm{cov}(x, y), \tag{1.25}$$

where

$$\mathrm{cov}(x, y) = \langle xy \rangle - \langle x \rangle \langle y \rangle \tag{1.26}$$

is the *covariance* of x and y, which can be positive or negative. A related quantity is the *correlation coefficient*,

$$\rho(x, y) = \frac{\mathrm{cov}(x, y)}{\sqrt{\mathrm{var}(x)\, \mathrm{var}(y)}}, \tag{1.27}$$

which takes values from -1 to 1. Notice that $\mathrm{cov}(x,x) = \mathrm{var}(x)$. When the variables x and y are independent, i.e. when $p(x,y) = p_x(x)\,p_y(y)$, we have

$$\mathrm{cov}(x,y) = 0 \quad \text{and} \quad \mathrm{var}(x+y) = \mathrm{var}(x) + \mathrm{var}(y). \tag{1.28}$$

Moreover, for independent variables,

$$\mathrm{var}\{a_1 x + a_2 y\} = a_1^2\,\mathrm{var}(x) + a_2^2\,\mathrm{var}(y). \tag{1.29}$$

1.2 Random sampling methods

The first component of a Monte Carlo calculation is the numerical sampling of random variables with specified PDFs. In this section we describe different techniques to generate random values of a variable x distributed in the interval $(x_{\min}, x_{\max})$ according to a given PDF $p(x)$. We concentrate on the simple case of single-variable distributions, since random sampling from multivariate distributions can always be reduced to single-variable sampling (see below). A more detailed description of sampling methods can be found in the textbooks of Rubinstein (1981) and Kalos and Whitlock (1986).

1.2.1 Random number generator

In general, random sampling algorithms are based on the use of random numbers ξ uniformly distributed in the interval (0,1). These random numbers can be easily generated on the computer (see e.g. Kalos and Whitlock, 1986; James, 1990). Among the "good" random number generators currently available, the simplest ones are the so-called multiplicative congruential generators (Press and Teukolsky, 1992). A popular example of this kind of generator is the following,

$$R_n = 7^5 R_{n-1} \pmod{2^{31}-1}, \qquad \xi_n = R_n/(2^{31}-1), \tag{1.30}$$

which produces a sequence of random numbers ξ_n uniformly distributed in (0,1) from a given "seed" R_0 ($< 2^{31}-1$). Actually, the generated sequence is not truly random, since it is obtained from a deterministic algorithm (the term "pseudo-random" would be more appropriate), but it is very unlikely that the subtle correlations between the values in the sequence have an appreciable effect on the simulation results. The generator (1.30) is known to have good random properties (Press and Teukolsky, 1992). However, the sequence is periodic, with a period of the order of 10^9. With present-day computational facilities, this value is not large enough to prevent re-initiation in a single simulation run. An excellent critical review of random number generators has been published by James (1990), where he recommends using algorithms that are more sophisticated than simple congruential ones. The generator implemented in the FORTRAN77 function RAND (table 1.1) is due to L'Ecuyer (1988); it produces 32-bit floating point numbers uniformly distributed in the *open* interval between zero and one. Its period is of the order of 10^{18}, which is virtually infinite for practical simulations.

Table 1.1: FORTRAN77 random number generator.

```
C  ***************************************************************************
C                          FUNCTION RAND
C  ***************************************************************************
      FUNCTION RAND(DUMMY)
C
C  This is an adapted version of subroutine RANECU written by F. James
C  (Comput. Phys. Commun. 60 (1990) 329-344), which has been modified to
C  give a single random number at each call.
C
C  The 'seeds' ISEED1 and ISEED2 must be initialized in the main program
C  and transferred through the named common block /RSEED/.
C
      IMPLICIT DOUBLE PRECISION (A-H,O-Z), INTEGER*4 (I)
      PARAMETER (USCALE=1.0D0/2.0D0**31)
      COMMON/RSEED/ISEED1,ISEED2
C
      I1=ISEED1/53668
      ISEED1=40014*(ISEED1-I1*53668)-I1*12211
      IF(ISEED1.LT.0) ISEED1=ISEED1+2147483563
C
      I2=ISEED2/52774
      ISEED2=40692*(ISEED2-I2*52774)-I2*3791
      IF(ISEED2.LT.0) ISEED2=ISEED2+2147483399
C
      IZ=ISEED1-ISEED2
      IF(IZ.LT.1) IZ=IZ+2147483562
      RAND=IZ*USCALE
C
      RETURN
      END
```

1.2.2 Inverse transform method

The cumulative distribution function of $p(x)$, eq. (1.6), is a non-decreasing function of x and, therefore, it has an inverse function $\mathcal{P}^{-1}(\xi)$. The transformation $\xi = \mathcal{P}(x)$ defines a new random variable that takes values in the interval (0,1), see fig. 1.1. Owing to the correspondence between x and ξ values, the PDF of ξ, $p_\xi(\xi)$, and that of x, $p(x)$, are related by $p_\xi(\xi)\,\mathrm{d}\xi = p(x)\,\mathrm{d}x$. Hence,

$$p_\xi(\xi) = p(x)\left(\frac{\mathrm{d}\xi}{\mathrm{d}x}\right)^{-1} = p(x)\left(\frac{\mathrm{d}\mathcal{P}(x)}{\mathrm{d}x}\right)^{-1} = 1, \tag{1.31}$$

that is, ξ is distributed uniformly in the interval (0,1).

Now it is clear that if ξ is a random number, the variable x defined by $x = \mathcal{P}^{-1}(\xi)$ is randomly distributed in the interval $(x_{\mathrm{min}}, x_{\mathrm{max}})$ with PDF $p(x)$ (see fig. 1.1). This provides a practical method of generating random values of x using a generator of random numbers uniformly distributed in (0,1). The randomness of x is guaranteed by

that of ξ. Notice that x is the (unique) root of the equation

$$\xi = \int_{x_{\min}}^{x} p(x')\,\mathrm{d}x', \tag{1.32}$$

which will be referred to as the *sampling equation* of the variable x. This procedure for random sampling is known as the *inverse transform method*; it is particularly adequate for PDFs $p(x)$ given by simple analytical expressions such that the sampling equation (1.32) can be solved analytically.

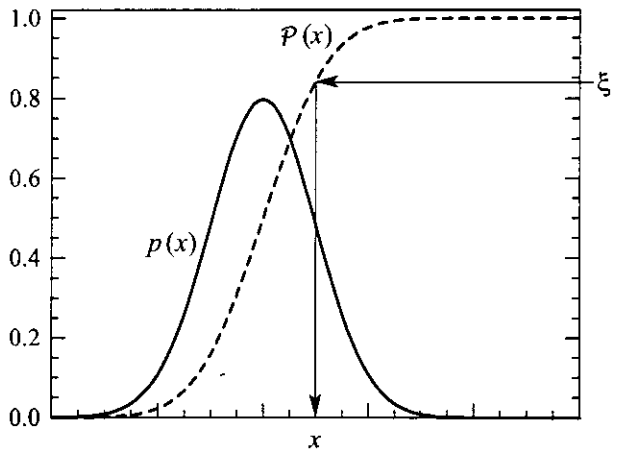


Figure 1.1: Random sampling from a distribution $p(x)$ using the inverse transform method.

Consider, for instance, the uniform distribution in the interval (a, b),

$$p(x) \equiv U_{a,b}(x) = \frac{1}{b-a}.$$

The sampling equation (1.32) then reads

$$\xi = \frac{x-a}{b-a}, \tag{1.33}$$

which leads to the well-known sampling formula

$$x = a + \xi(b-a). \tag{1.34}$$

As another familiar example, consider the exponential distribution

$$p(s) = \frac{1}{\lambda}\exp(-s/\lambda), \qquad s > 0, \tag{1.35}$$

of the free path s of a particle between interaction events (see section 1.4.1). The parameter λ represents the mean free path. In this case, the sampling equation (1.32) is easily solved to give the sampling formula

$$s = -\lambda \ln(1-\xi) = -\lambda \ln \xi. \tag{1.36}$$

The last equality follows from the fact that $1-\xi$ is also a random number distributed in (0,1).

Numerical inverse transform

The inverse transform method can also be efficiently used for random sampling from continuous distributions $p(x)$ that are given in numerical form, or that are too complicated to be sampled analytically. To apply this method, the cumulative distribution function $\mathcal{P}(x)$ has to be evaluated at the points x_i of a certain grid. The sampling equation $\mathcal{P}(x) = \xi$ can then be solved by inverse interpolation, i.e. by interpolating in the table (ξ_i, x_i), where $\xi_i \equiv \mathcal{P}(x_i)$ (ξ is regarded as the independent variable). Care must be exercised to make sure that the numerical integration and interpolation do not introduce significant errors.

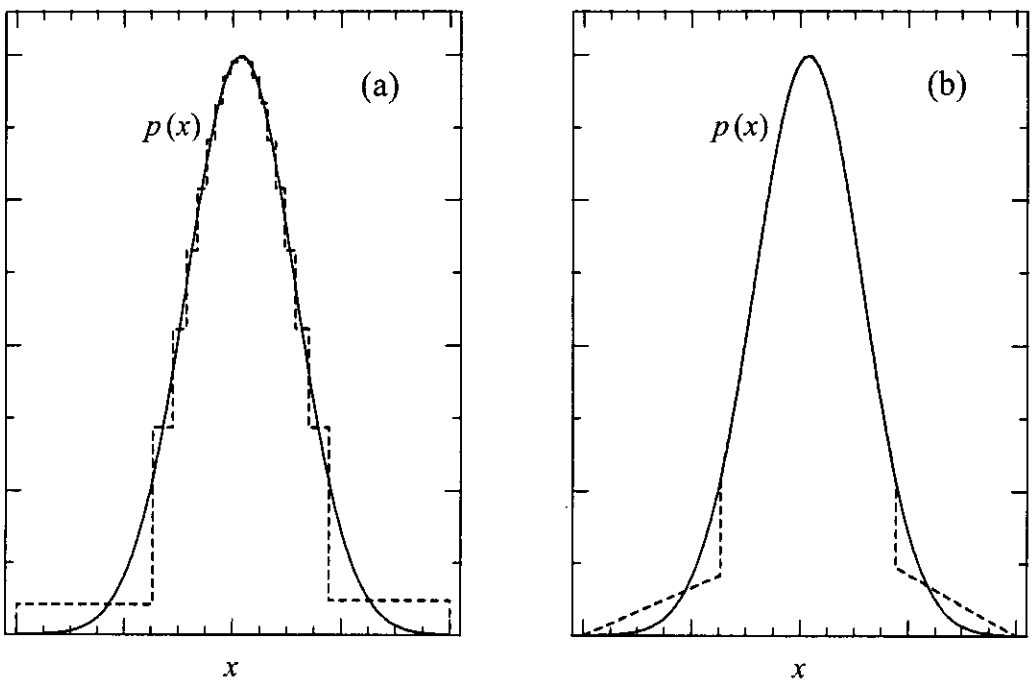


Figure 1.2: Random sampling from a continuous distribution $p(x)$ using the numerical inverse transform method with $N = 20$ grid points. a) Piecewise constant approximation. b) Piecewise linear approximation.

A simple, general, approximate method for numerical sampling from continuous distributions is the following. The values x_n ($n = 0, 1, \ldots, N$) of x for which the cumulative distribution function has the values n/N,

$$\mathcal{P}(x_n) = \int_{x_{\min}}^{x_n} p(x)\,\mathrm{d}x = \frac{n}{N}, \tag{1.37}$$

are previously computed and stored in memory. Notice that the *exact* probability of having x in the interval (x_n, x_{n+1}) is $1/N$. We can now sample x by linear interpolation:

 3444

we generate a random number ξ and consider the quantity $y \equiv \xi N$, which takes values in the interval $(0, N)$. We set $n = [y]$, where the symbol $[y]$ denotes the integer part of y (i.e. the largest integer that is less than y). The value of x is obtained as

$$x = x_n + (x_{n+1} - x_n)u, \qquad u \equiv y - n \in (0, 1). \tag{1.38}$$

This is equivalent to approximating the PDF by a piecewise constant function (see fig. 1.2a). Since the spacing between the points x_n (at which the cumulative distribution function is specified) is roughly proportional to $1/p(x_n)$, the approximation is more accurate in regions where $p(x)$ is large.

The algorithm can be improved by storing the values $p(x_n)$ of the PDF at the points x_n in memory and approximating the PDF in the interval (x_n, x_{n+1}) linearly,

$$p_{\rm la}(x) \simeq C_n \left[p(x_n) + \frac{p(x_{n+1}) - p(x_n)}{x_{n+1} - x_n} (x - x_n) \right], \tag{1.39}$$

with a normalization constant C_n such that the integral of $p_{\rm la}(x)$ over the interval (x_n, x_{n+1}) equals $1/N$. In general, this piecewise linear approximation is not continuous. Of course, $p_{\rm la}(x)$ will differ from the exact PDF $p(x)$ when the latter is not linear in the interval, but the differences are smaller than for the piecewise constant approximation with the same number N of grid points (see fig. 1.2). Again, the approximation is better where $p(x)$ is larger. An exact algorithm for random sampling from the piecewise linear approximation (1.39) is the following,

(i) Generate a random number ξ and set $y = \xi N$, $n = [y]$ and $u = y - n$.
(ii) If $p(x_n) \neq 0$, set $r = p(x_{n+1})/p(x_n)$ and

$$t = \begin{cases} \dfrac{(1 - u + r^2 u)^{1/2} - 1}{r - 1} & \text{if } r \neq 1, \\ u & \text{if } r = 1. \end{cases} \tag{1.40}$$

(iii) If $p(x_n) = 0$, set $t = u^{1/2}$.
(iv) Deliver $x = x_n + (x_{n+1} - x_n)t$.

1.2.3 Discrete distributions

The inverse transform method can also be applied to discrete distributions. Consider that the random variable x can take the discrete values $x = 1, \ldots, N$ with point probabilities $p_1, \ldots, p_N$, respectively. The corresponding PDF can be expressed as

$$p(x) = \sum_{i=1}^{N} p_i \delta(x - i), \tag{1.41}$$

where $\delta(x)$ is the Dirac distribution. Here $p(x)$ is assumed to be defined in an interval (a, b) with $a < 1$ and $b > N$. The corresponding cumulative distribution function is

$$\mathcal{P}(x) = \sum_{i=1}^{[x]} p_i. \tag{1.42}$$

where $[x]$ stands for the integer part of x. Notice that $\mathcal{P}(x) = 0$ when $x < 1$. Then, eq. (1.32) leads to the sampling formula

$$\begin{aligned} x &= 1 \quad \text{if } \xi \leq p_1 \\ &= 2 \quad \text{if } p_1 < \xi \leq p_1 + p_2 \\ &\vdots \\ &= j \quad \text{if } \textstyle\sum_{i=1}^{j-1} p_i < \xi \leq \sum_{i=1}^{j} p_i \\ &\vdots \end{aligned} \tag{1.43}$$

We can define the quantities

$$P_1 = 0, \quad P_2 = p_1, \quad P_3 = p_1 + p_2, \quad \ldots, \quad P_{N+1} = \sum_{i=1}^{N} p_i = 1. \tag{1.44}$$

To sample x we generate a random number ξ and set x equal to the index i such that

$$P_i < \xi \leq P_{i+1}. \tag{1.45}$$

If the number N of x-values is large, this sampling algorithm may be quite slow because of the large number of comparisons needed to determine the sampled value. The easiest method to reduce the number of comparisons is to use binary search instead of sequential search. The algorithm for binary search, for a given value of ξ, proceeds as follows:

(i) Set $i = 1$ and $j = N + 1$.
(ii) Set $k = [(i + j)/2]$.
(iii) If $P_k < \xi$, set $i = k$; otherwise set $j = k$.
(iv) If $j - i > 1$, go to step (ii).
(v) Deliver i.

When $2^n < N \leq 2^{n+1}$, i is obtained after $n+1$ comparisons. This number of comparisons is evidently much less than the number required when using purely sequential search.

Walker's aliasing method

Walker (1977) described an optimal sampling method for discrete distributions, which yields the sampled value with only one comparison. The idea underlying Walker's method can be easily understood by resorting to graphical arguments (Salvat, 1987). For this purpose, let us represent the PDF (1.41) as a histogram constructed with N bars of width $1/N$ and heights Np_i (see fig. 1.3). Now, the histogram bars can be cut off at convenient heights and the resulting pieces can be arranged to fill up the square of unit side in such a way that each vertical line crosses, at most, two different pieces. This arrangement can be performed systematically by selecting the lowest and the highest bars in the histogram, say the ℓth and the jth, respectively, and by cutting the highest bar off to complete the lowest one, which subsequently is kept unaltered. In order to

keep track of the performed transformation, we label the added piece with the "alias" value $K_\ell = j$, giving its original position in the histogram, and introduce the "cutoff" value F_ℓ defined as the height of the lower piece in the ℓth bar of the resulting square. This lower piece keeps the label ℓ. Evidently, iteration of this process eventually leads to the complete square (after $N-1$ steps). Notice that the point probabilities p_i can be reconstructed from the alias and cutoff values. We have

$$Np_i = F_i + \sum_{j \neq i}(1 - F_j)\delta(i, K_j), \tag{1.46}$$

where $\delta(i,j)$ denotes the Kronecker delta (= 1 if $i = j$ and = 0 otherwise). Walker's method for random sampling of x proceeds as follows: We sample two independent random numbers, say ξ_1 and ξ_2, and define the random point (ξ_1,ξ_2), which is uniformly distributed in the square. If (ξ_1,ξ_2) lies over a piece labelled with the index i, we take $x = i$ as the selected value. Obviously, the probability of obtaining i as a result of the sampling equals the fractional area of the pieces labelled with i, which coincides with p_i.

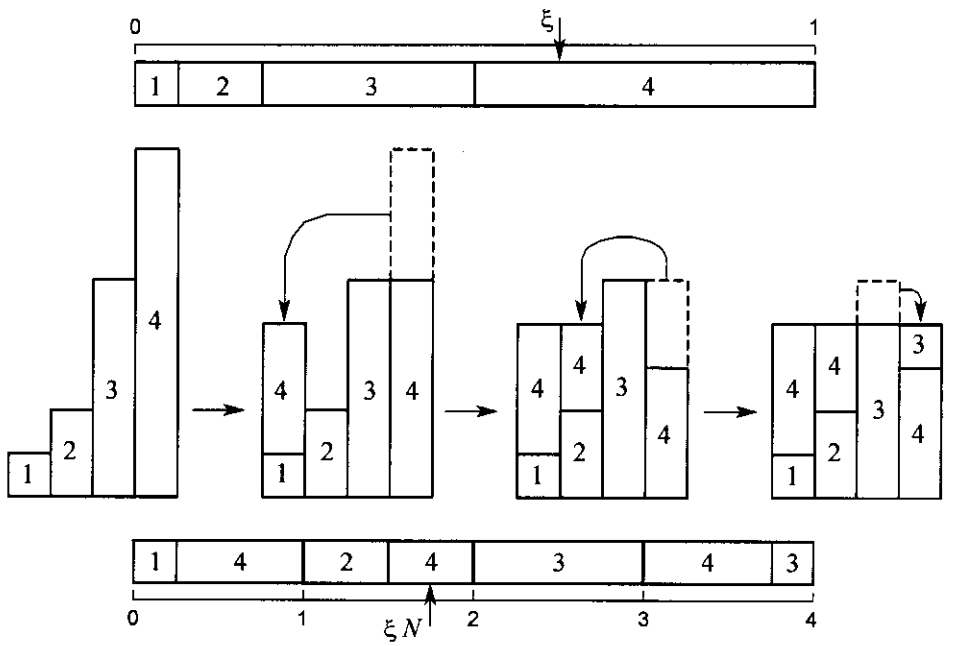


Figure 1.3: Graphical representation of the inverse transform method (top) and Walker's aliasing method (bottom) for random sampling from a discrete distribution. In this example, the random variable can take the values $i = 1$, 2, 3 and 4 with relative probabilities 1, 2, 5 and 8, respectively.

As formulated above, Walker's algorithm requires the generation of two random numbers for each sampled value of x. With the aid of the following trick, the x-value

can be generated from a single random number. Continuing with our graphical picture, assume that the N bars in the square are aligned consecutively to form a segment of length N (bottom of fig. 1.3). To sample x, we can generate a single random value ξN, which is uniformly distributed in $(0,N)$ and determines one of the segment pieces. The result of the sampling is the label of the selected piece. Explicitly, the sampling algorithm proceeds as follows:

(i) Generate a random number ξ and set $R = \xi N + 1$.
(ii) Set $i = [R]$ and $r = R - i$.
(iii) If $r > F_i$, deliver $x = K_i$.
(iv) Deliver $x = i$.

We see that the sampling of x involves only the generation of a random number and one comparison (irrespective of the number N of possible outcomes). The price we pay for this simplification reduces to doubling the number of memory locations that are needed: the two arrays K_i and F_i are used instead of the single array p_i (or P_i). Unfortunately, the calculation of alias and cutoff values is fairly involved and this limits the applicability of Walker's algorithm to distributions that remain constant during the course of the simulation.

1.2.4 Rejection methods

The inverse transform method for random sampling is based on a one-to-one correspondence between x and ξ values, which is expressed in terms of a single-valued function. There is another kind of sampling method, due to von Neumann, that consists of sampling a random variable from a certain distribution [different from $p(x)$] and subjecting it to a random test to determine whether it will be accepted for use or rejected. These rejection methods lead to very general techniques for sampling from any PDF.

The rejection algorithms can be understood in terms of simple graphical arguments (fig. 1.4). Consider that, by means of the inverse transform method or any other available sampling method, random values of x are generated from a PDF $\pi(x)$. For each sampled value of x we sample a random value y uniformly distributed in the interval $(0, C\pi(x))$, where C is a positive constant. Evidently, the points (x, y), generated in this way, are uniformly distributed in the region A of the plane limited by the x-axis ($y = 0$) and the curve $y = C\pi(x)$. Conversely, if (by some means) we generate random points (x, y) uniformly distributed in A, their x-coordinate is a random variable distributed according to $\pi(x)$ (irrespective of the value of C). Now, consider that the distribution $\pi(x)$ is such that $C\pi(x) \geq p(x)$ for some $C > 0$ and that we generate random points (x, y) uniformly distributed in the region A as described above. If we reject the points with $y > p(x)$, the accepted ones (with $y \leq p(x)$) are uniformly distributed in the region between the x-axis and the curve $y = p(x)$ and hence, their x-coordinate is distributed according to $p(x)$.

A rejection method is thus completely specified by representing the PDF $p(x)$ as

$$p(x) = C\pi(x)r(x), \tag{1.47}$$

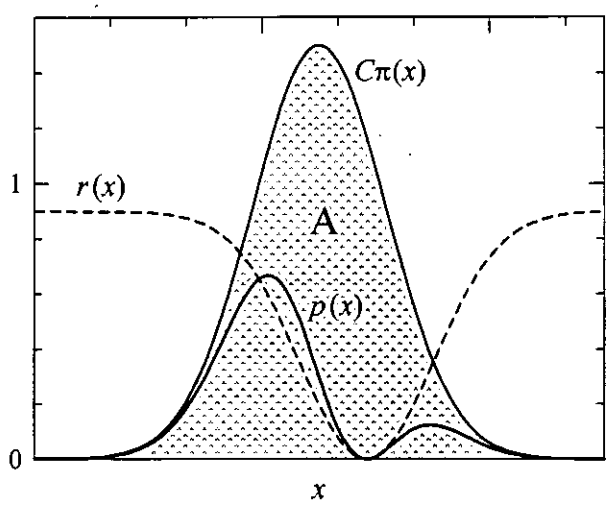


Figure 1.4: Random sampling from a distribution $p(x)$ using a rejection method.

where $\pi(x)$ is a PDF that can be easily sampled e.g. by the inverse transform method, C is a positive constant and the function $r(x)$ satisfies the conditions $0 < r(x) \leq 1$. The rejection algorithm for sampling from $p(x)$ proceeds as follows:

(i) Generate a random value x from $\pi(x)$.
(ii) Generate a random number ξ.
(iii) If $\xi > r(x)$, go to step (i).
(iv) Deliver x.

From the geometrical arguments given above, it is clear that the algorithm does yield x values distributed according to $p(x)$. The following is a more formal proof: Step (i) produces x-values in the interval $(x, x + \mathrm{d}x)$ with probability $\pi(x)\,\mathrm{d}x$, these values are accepted with probability $r(x) = p(x)/[C\pi(x)]$ and, therefore, (apart from a normalization constant) the probability of delivering a value in $(x, x + \mathrm{d}x)$ is equal to $p(x)\,\mathrm{d}x$ as required. It is important to realize that, as regards Monte Carlo, the normalization of the simulated PDF is guaranteed by the mere fact that the algorithm delivers some value of x.

The efficiency of the algorithm, i.e. the probability of accepting a generated x-value, is

$$\epsilon = \int_a^b r(x)\pi(x)\,\mathrm{d}x = \frac{1}{C}. \tag{1.48}$$

Graphically, the efficiency equals the ratio of the areas under the curves $y = p(x)$ and $y = C\pi(x)$, which are 1 and C, respectively. For a given $\pi(x)$, since $r(x) \leq 1$, the constant C must satisfy the condition $C\pi(x) \geq p(x)$ for all x. The minimum value of C, with the requirement that $C\pi(x) = p(x)$ for some x, gives the optimum efficiency.

The PDF $\pi(x)$ in eq. (1.47) should be selected in such a way that the resulting sampling algorithm is as fast as possible. In particular, random sampling from $\pi(x)$

must be performed rapidly, by the inverse transform method or by the composition method (see below). High efficiency is also desirable, but not decisive. One hundred percent efficiency is obtained only with $\pi(x) = p(x)$ (but random sampling from this PDF is just the problem we want to solve); any other PDF gives a lower efficiency. The usefulness of the rejection method lies in the fact that a certain loss of efficiency can be largely compensated with the ease of sampling x from $\pi(x)$ instead of $p(x)$. A disadvantage of this method is that it requires the generation of several random numbers ξ to sample each x-value.

1.2.5 Two-dimensional variables. Composition methods

Let us consider a two-dimensional random variable (x, y) with joint probability distribution $p(x, y)$. Introducing the marginal PDF $q(y)$ and the conditional PDF $p(x|y)$ [see eqs. (1.18) and (1.20)],

$$q(y) \equiv \int p(x, y)\, \mathrm{d}x, \qquad p(x|y) = \frac{p(x, y)}{q(y)},$$

the two-variate distribution can be expressed as

$$p(x, y) = q(y)\, p(x|y). \tag{1.49}$$

It is now evident that to generate random points (x, y) from $p(x, y)$ we can first sample y from $q(y)$ and then x from $p(x|y)$. Hence, two-dimensional random variables can be generated by using single-variable sampling methods. This is also true for multivariate distributions, because an n-dimensional PDF can always be expressed as the product of a single-variable marginal distribution and an $(n-1)$-dimensional conditional PDF.

From the definition of the marginal PDF of x,

$$q(x) \equiv \int p(x, y)\, \mathrm{d}y = \int q(y)\, p(x|y)\, \mathrm{d}y, \tag{1.50}$$

it is clear that if we sample y from $q(y)$ and, then, x from $p(x|y)$, the generated values of x are distributed according to $q(x)$. This idea is the basis of the *composition* methods, which are applicable when $p(x)$, the distribution to be simulated, is a probability mixture of several PDFs. More specifically, we consider that $p(x)$ can be expressed as

$$p(x) = \int w(y)\, p_y(x)\, \mathrm{d}y, \tag{1.51}$$

where $w(y)$ is a continuous distribution and $p_y(x)$ is a family of one-parameter PDFs, where y is the parameter identifying a unique distribution. Notice that if the parameter y takes only integer values $y = i$ with point probabilities w_i, we would write

$$p(x) = \sum_i w_i\, p_i(x). \tag{1.52}$$

 3450

The composition method for random sampling from the PDF $p(x)$ is as follows. First, a value of y (or i) is drawn from the PDF $w(y)$ and then x is sampled from the PDF $p_y(x)$ for that chosen y.

This technique may be applied to generate random values from complex distributions obtained by combining simpler distributions that are themselves easily generated, by the inverse transform method or by rejection methods.

Devising fast, exact methods for random sampling from a given PDF is an interesting technical challenge. The ultimate criterion for the quality of a sampling algorithm is its speed in actual simulations: the best algorithm is the fastest. However, programming simplicity and elegance may justify the use of slower algorithms. For simple analytical distributions that have an analytical inverse cumulative distribution function, the inverse transform method is usually satisfactory. This is the case for a few elementary distributions (e.g. the uniform and exponential distributions considered above). The inverse transform method is also adequate for discrete distributions and for continuous PDFs given in numerical form. By combining the inverse transform, rejection and composition methods we can devise sampling algorithms for virtually any (single- or multivariate) PDF.

Example 1. Sampling from the normal distribution

Frequently, we need to generate random values from the normal (or Gaussian) distribution

$$p_{\mathrm{G}}(x) = \frac{1}{\sqrt{2\pi}} \exp(-x^2/2). \tag{1.53}$$

Since the cumulative distribution function cannot be inverted analytically, the inverse transform method is not appropriate. The easiest (but not the fastest) method to sample from the normal distribution consists of generating two independent random variables at a time, as follows. Let x_1 and x_2 be two independent normal variables. They determine a random point in the plane with PDF

$$p_{2\mathrm{G}}(x_1, x_2) = p_{\mathrm{G}}(x_1)\, p_{\mathrm{G}}(x_2) = \frac{1}{2\pi} \exp[-(x_1^2 + x_2^2)/2].$$

Introducing the polar coordinates r and ϕ,

$$x_1 = r\cos\phi, \qquad x_2 = r\sin\phi,$$

the PDF can be expressed as

$$p_{2\mathrm{G}}(x_1, x_2)\, \mathrm{d}x_1\, \mathrm{d}x_2 = \frac{1}{2\pi} \exp(-r^2/2)\, r\mathrm{d}r\, \mathrm{d}\phi = \left[\exp(-r^2/2)\, r\mathrm{d}r\right] \left[\frac{1}{2\pi}\mathrm{d}\phi\right].$$

We see that r and ϕ are independent random variables. The angle ϕ is distributed uniformly on $(0,2\pi)$ and can be sampled as $\phi = 2\pi\xi$. The PDF of r is $\exp(-r^2/2)\, r$ and

the corresponding cumulative distribution function is $\mathcal{P}(r) = 1-\exp(-r^2/2)$. Therefore, r can be generated by the inverse transform method as

$$r = \sqrt{-2\ln(1-\xi)} = \sqrt{-2\ln\xi}.$$

The two independent normal random variables are given by

$$\begin{aligned} x_1 &= \sqrt{-2\ln\xi_1}\,\cos(2\pi\xi_2), \\ x_2 &= \sqrt{-2\ln\xi_1}\,\sin(2\pi\xi_2), \end{aligned} \tag{1.54}$$

where ξ_1 and ξ_2 are two independent random numbers. This procedure is known as the Box-Müller method. It has the advantages of being exact and easy to program (it can be coded as a single FORTRAN statement).

The mean and variance of the normal variable are $\langle x\rangle = 0$ and $\mathrm{var}(x) = 1$. The linear transformation

$$X = m + \sigma x \qquad (\sigma > 0) \tag{1.55}$$

defines a new random variable. From the properties (1.14) and (1.29), we have

$$\langle X\rangle = m \quad \text{and} \quad \mathrm{var}(X) = \sigma^2. \tag{1.56}$$

The PDF of X is

$$p(X) = p_{\mathrm{G}}(x)\frac{\mathrm{d}x}{\mathrm{d}X} = \frac{1}{\sigma\sqrt{2\pi}}\exp\left[-\frac{(X-m)^2}{2\sigma^2}\right], \tag{1.57}$$

i.e. X is normally distributed with mean m and variance σ^2. Hence, to generate X we only have to sample x using the Box-Müller method and apply the transformation (1.55).

Example 2. Uniform distribution on the unit sphere

In radiation transport, the direction of motion of a particle is described by a unit vector $\hat{\mathbf{d}}$. Given a certain frame of reference, the direction $\hat{\mathbf{d}}$ can be specified by giving either its direction cosines (u, v, w) (i.e. the projections of $\hat{\mathbf{d}}$ on the directions of the coordinate axes) or the polar angle θ and the azimuthal angle ϕ, defined as in fig. 1.5,

$$\hat{\mathbf{d}} = (u, v, w) = (\sin\theta\,\cos\phi, \sin\theta\,\sin\phi, \cos\theta). \tag{1.58}$$

Notice that $\theta \in (0, \pi)$ and $\phi \in (0, 2\pi)$.

A direction vector can be regarded as a point on the surface of the unit sphere. Consider an isotropic source of particles, i.e. such that the initial direction (θ, ϕ) of emitted particles is a random point uniformly distributed on the surface of the sphere. The PDF is

$$p(\theta,\phi)\,\mathrm{d}\theta\,\mathrm{d}\phi = \frac{1}{4\pi}\sin\theta\,\mathrm{d}\theta\,\mathrm{d}\phi = \left[\frac{\sin\theta}{2}\mathrm{d}\theta\right]\left[\frac{1}{2\pi}\mathrm{d}\phi\right]. \tag{1.59}$$

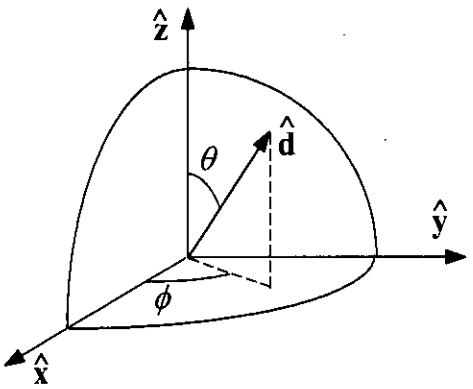


Figure 1.5: Polar and azimuthal angles of a direction vector.

That is, θ and ϕ are independent random variables with PDFs $p_\theta(\theta) = \sin\theta/2$ and $p_\phi(\phi) = 1/(2\pi)$, respectively. Therefore, the initial direction of a particle from an isotropic source can be generated by applying the inverse transform method to these PDFs,

$$\theta = \arccos(1 - 2\xi_1), \qquad \phi = 2\pi\xi_2. \tag{1.60}$$

In some cases, it is convenient to replace the polar angle θ by the variable

$$\mu = (1 - \cos\theta)/2, \tag{1.61}$$

which varies from 0 ($\theta = 0$) to 1 ($\theta = \pi$). In the case of an isotropic distribution, the PDF of μ is

$$p_\mu(\mu) = p_\theta(\theta)\left(\frac{\mathrm{d}\mu}{\mathrm{d}\theta}\right)^{-1} = 1. \tag{1.62}$$

That is, a set of random points (μ, ϕ) uniformly distributed on the rectangle $(0, 1) \times (0, 2\pi)$ corresponds to a set of random directions (θ, ϕ) uniformly distributed on the unit sphere.

1.3 Monte Carlo integration

As pointed out by James (1980), at least in a formal sense, all Monte Carlo calculations are equivalent to integrations. This equivalence permits a formal theoretical foundation for Monte Carlo techniques. An important aspect of simulation is the evaluation of the statistical uncertainties of the calculated quantities. We shall derive the basic formulae by considering the simplest Monte Carlo calculation, namely, the evaluation of a unidimensional integral. Evidently, the results are also valid for multidimensional integrals.

Consider the integral

$$I = \int_a^b F(x)\,\mathrm{d}x, \tag{1.63}$$

which we recast in the form of an expectation value,

$$I = \int f(x)\,p(x)\,dx \equiv \langle f \rangle, \tag{1.64}$$

by introducing an arbitrary PDF $p(x)$ and setting $f(x) = F(x)/p(x)$ [it is assumed that $p(x) > 0$ in (a, b) and $p(x) = 0$ outside this interval]. The Monte Carlo evaluation of the integral I is very simple: generate a large number N of random points x_i from the PDF $p(x)$ and accumulate the sum of values $f(x_i)$ in a counter. At the end of the calculation the expected value of f is estimated as

$$\overline{f} \equiv \frac{1}{N}\sum_{i=1}^{N} f(x_i). \tag{1.65}$$

The law of large numbers says that, as N becomes very large,

$$\overline{f} \to I \quad \text{(in probability)}. \tag{1.66}$$

In statistical terminology, this means that $\overline{f}$, the Monte Carlo result, is a *consistent estimator* of the integral (1.63). This is valid for any function $f(x)$ that is finite and piecewise continuous, i.e. with a finite number of discontinuities.

The law of large numbers (1.66) can be restated as

$$\langle f \rangle = \lim_{N\to\infty} \frac{1}{N}\sum_{i=1}^{N} f(x_i). \tag{1.67}$$

By applying this law to the integral that defines the variance of $f(x)$ [cf. eq. (1.16)]

$$\mathrm{var}\{f(x)\} = \int f^2(x)\,p(x)\,\mathrm{d}x - \langle f \rangle^2, \tag{1.68}$$

we obtain

$$\mathrm{var}\{f(x)\} = \lim_{N\to\infty} \left\{ \frac{1}{N}\sum_{i=1}^{N} [f(x_i)]^2 - \left[\frac{1}{N}\sum_{i=1}^{N} f(x_i)\right]^2 \right\}. \tag{1.69}$$

The expression in curly brackets is a consistent estimator of the variance of $f(x)$. It is advisable (see below) to accumulate the squared function values $[f(x_i)]^2$ in a counter and, at the end of the simulation, estimate $\mathrm{var}\{f(x)\}$ according to eq. (1.69).

It is clear that different Monte Carlo runs [with different, independent sequences of N random numbers x_i from $p(x)$] will yield different estimates $\overline{f}$. This implies that the outcome of our Monte Carlo code is affected by statistical uncertainties, similar to those found in laboratory experiments, which need to be properly evaluated to determine the "accuracy" of the Monte Carlo result. For this purpose, we may consider $\overline{f}$ as a random

variable, the PDF of which is, in principle, unknown. Its mean and variance are given by

$$\langle \overline{f} \rangle = \left\langle \frac{1}{N} \sum_{i=1}^{N} f(x_i) \right\rangle = \frac{1}{N} \sum_{i=1}^{N} \langle f \rangle = \langle f \rangle \tag{1.70}$$

and

$$\mathrm{var}(\overline{f}) = \mathrm{var}\left[\frac{1}{N} \sum_{i=1}^{N} f(x_i) \right] = \frac{1}{N^2} \sum_{i=1}^{N} \mathrm{var}\{f(x)\} = \frac{1}{N} \mathrm{var}\{f(x)\}, \tag{1.71}$$

where use has been made of properties of the expectation and variance operators. The standard deviation (or standard error) of $\overline{f}$,

$$\sigma_f \equiv \sqrt{\mathrm{var}(\overline{f})} = \sqrt{\frac{\mathrm{var}\{f(x)\}}{N}}, \tag{1.72}$$

gives a measure of the statistical uncertainty of the Monte Carlo estimate $\overline{f}$. The result (1.72) has an important practical implication: in order to reduce the statistical uncertainty by a factor of 10, we have to increase the sample size N by a factor of 100. Evidently, this sets a limit to the accuracy that can be attained with the available computer power.

We can now invoke the central limit theorem (see e.g. James, 1980), which establishes that, in the limit $N \to \infty$, the PDF of $\overline{f}$ is a normal (Gaussian) distribution with mean $\langle f \rangle$ and standard deviation σ_f,

$$p(\overline{f}) = \frac{1}{\sigma_f \sqrt{2\pi}} \exp\left(-\frac{(\overline{f} - \langle f \rangle)^2}{2\sigma_f^2} \right). \tag{1.73}$$

It follows that, for sufficiently large values of N, for which the theorem is applicable, the interval $\overline{f} \pm n\sigma_f$ contains the exact value $\langle f \rangle$ with a probability of 68.3% if $n = 1$, 95.4% if $n = 2$ and 99.7% if $n = 3$ (3σ rule).

The central limit theorem is a very powerful tool, since it predicts that the generated values of $\overline{f}$ follow a specific distribution, but it applies only asymptotically. The minimum number N of sampled values needed to apply the theorem with confidence depends on the problem under consideration. If, in the case of our problem, the third central moment of f,

$$\mu_3 \equiv \int [f(x) - \langle f \rangle]^3 \, p(x) \, \mathrm{d}x, \tag{1.74}$$

exists, the theorem is essentially satisfied when

$$|\mu_3| \ll \sigma_f^3 \sqrt{N}. \tag{1.75}$$

In general, it is advisable to study the distribution of the estimator to ascertain the applicability of the central limit theorem. In most Monte Carlo calculations, however, statistical errors are estimated by simply assuming that the theorem is satisfied, irrespective of the sample size. We shall adopt this practice and report Monte Carlo results in the form $\overline{f} \pm 3\sigma_f$. In simulations of radiation transport, this is empirically validated

by the fact that simulated continuous distributions do "look" continuous (i.e. the "error bars" define a smooth band).

Each possible $p(x)$ defines a Monte Carlo algorithm to calculate the integral I, eq. (1.63). The simplest algorithm (crude Monte Carlo) is obtained by using the uniform distribution $p(x) = 1/(b-a)$. Evidently, $p(x)$ determines not only the density of sampled points x_i, but also the magnitude of the variance $\mathrm{var}\{f(x)\}$, eq. (1.68),

$$\mathrm{var}\{f(x)\} = \int_a^b p(x) \left[\frac{F(x)}{p(x)}\right]^2 \mathrm{d}x - I^2 = \int_a^b F(x) \left[\frac{F(x)}{p(x)} - I\right] \mathrm{d}x. \tag{1.76}$$

As a measure of the effectiveness of a Monte Carlo algorithm, it is common to use the efficiency ϵ, which is defined by

$$\epsilon = 1/[\sigma_f^2 T], \tag{1.77}$$

where T is the computing time (or any other measure of the calculation effort) needed to get the simulation result. Since σ_f^2 and T are roughly proportional to N^{-1} and N, respectively, ϵ is a constant (i.e. it is independent of N), on average.

The so-called variance-reduction methods are techniques that aim to optimize the *efficiency* of the simulation through an adequate choice of the PDF $p(x)$. Improving the efficiency of the algorithms is an important, and delicate, part of the art of Monte Carlo simulation. The interested reader is addressed to the specialized bibliography (e.g. Rubinstein, 1981). Although of common use, the term "variance reduction" is somewhat misleading, since a reduction in variance does not necessarily lead to improved efficiency. To make this clear, consider that a Monte Carlo algorithm, based on a certain PDF $p(x)$, has a variance that is less than that of crude Monte Carlo (i.e. with the uniform distribution); if the generation of x-values from $p(x)$ takes a longer time than for the uniform distribution, the "variance-reduced" algorithm may be less efficient than crude Monte Carlo. Hence, one should avoid using PDFs that are too difficult to sample.

1.4 Simulation of radiation transport

In this section, we describe the essentials of Monte Carlo simulation of radiation transport. For the sake of simplicity, we limit our considerations to the detailed simulation method, where all the interaction events experienced by a particle are simulated in chronological succession, and we disregard the production of secondary particles, so that only one kind of particle is transported.

1.4.1 Scattering model and probability distribution functions

Consider a particle with energy E (kinetic energy, in the case of electrons and positrons) moving in a given medium. We limit our considerations to homogeneous "random scattering" media, such as gases, liquids and amorphous solids, where the "molecules"

are distributed at random with uniform density. The composition of the medium is specified by its stoichiometric formula, i.e. atomic number Z_i and number of atoms per molecule n_i of all the elements present. The stoichiometric indices n_i need not have integer values. In the case of alloys, for instance, they may be set equal to the percentage in number of each element and then a "molecule" is a group of 100 atoms with the appropriate proportion of each element. The "molecular weight" is $A_{\mathrm{M}} = \Sigma n_i A_i$, where A_i is the atomic weight of the i-th element. The number of molecules per unit volume is given by

$$\mathcal{N} = N_{\mathrm{A}} \frac{\rho}{A_{\mathrm{M}}}, \tag{1.78}$$

where N_{A} is Avogadro's number and ρ is the mass density of the material.

In each interaction, the particle may lose energy W and/or change its direction of movement. The angular deflection is determined by the polar scattering angle θ, i.e. the angle between the directions of the particle before and after the interaction, and the azimuthal angle ϕ. Let us assume that the particle can interact with the medium through two independent mechanisms, denoted as "A" and "B" (for instance, elastic and inelastic scattering, in the case of low-energy electrons). The scattering model is completely specified by the molecular differential cross sections (DCS)

$$\frac{\mathrm{d}^2\sigma_{\mathrm{A}}}{\mathrm{d}W\mathrm{d}\Omega}(E;W,\theta) \qquad \text{and} \qquad \frac{\mathrm{d}^2\sigma_{\mathrm{B}}}{\mathrm{d}W\mathrm{d}\Omega}(E;W,\theta), \tag{1.79}$$

where $\mathrm{d}\Omega$ is a solid angle element in the direction (θ,ϕ). We have made the parametric dependence of the DCSs on the particle energy E explicit. Considering that the molecules in the medium are oriented at random, the DCS is independent of the azimuthal scattering angle, i.e. the angular distribution of scattered particles is axially symmetrical around the direction of incidence. The total cross sections (per molecule) are

$$\sigma_{\mathrm{A,B}}(E) = \int_0^E \mathrm{d}W \int_0^\pi 2\pi \sin\theta \,\mathrm{d}\theta \, \frac{\mathrm{d}^2\sigma_{\mathrm{A,B}}}{\mathrm{d}W\mathrm{d}\Omega}(E;W,\theta). \tag{1.80}$$

The PDFs of the energy loss and the polar scattering angle in individual scattering events are

$$p_{\mathrm{A,B}}(E;W,\theta) = \frac{2\pi \sin\theta}{\sigma_{\mathrm{A,B}}(E)} \frac{\mathrm{d}^2\sigma_{\mathrm{A,B}}}{\mathrm{d}W\mathrm{d}\Omega}(E;W,\theta). \tag{1.81}$$

Notice that $p_{\mathrm{A}}(E;W,\theta)\mathrm{d}W\mathrm{d}\theta$ gives the (normalized) probability that, in a scattering event of type A, the particle loses energy in the interval $(W, W+\mathrm{d}W)$ and is deflected into directions with polar angle (relative to the initial direction) in the interval $(\theta, \theta+\mathrm{d}\theta)$. The azimuthal scattering angle in each collision is uniformly distributed in the interval $(0, 2\pi)$, i.e.

$$p(\phi) = \frac{1}{2\pi}. \tag{1.82}$$

The total interaction cross section is

$$\sigma_{\mathrm{T}}(E) = \sigma_{\mathrm{A}}(E) + \sigma_{\mathrm{B}}(E). \tag{1.83}$$

When the particle interacts with the medium, the kind of interaction that occurs is a discrete random variable, that takes the values "A" and "B" with probabilities

$$p_{\mathrm{A}} = \sigma_{\mathrm{A}}/\sigma_{\mathrm{T}} \qquad \text{and} \qquad p_{\mathrm{B}} = \sigma_{\mathrm{B}}/\sigma_{\mathrm{T}}. \tag{1.84}$$

It is worth recalling that this kind of single scattering model is only valid when diffraction effects resulting from coherent scattering from several centres (e.g. Bragg diffraction, channelling of charged particles) are negligible. This means that the simulation is applicable only to amorphous media and, with some care, to polycrystalline solids.

To get an intuitive picture of the scattering process, we can imagine each molecule as a sphere of radius r_{s} such that the cross-sectional area πr_{s}^2 equals the total cross section σ_{T}. Now, assume that a particle impinges normally on a very thin material foil of thickness $\mathrm{d}s$. What the particle sees in front of it is a uniform distribution of $\mathcal{N}\,\mathrm{d}s$ spheres per unit surface. An interaction takes place when the particle strikes one of these spheres. Therefore, the probability of interaction within the foil equals the fractional area covered by the spheres, $\mathcal{N}\sigma_{\mathrm{T}}\,\mathrm{d}s$. In other words, $\mathcal{N}\sigma_{\mathrm{T}}$ is the interaction probability per unit path length. Its inverse,

$$\lambda_{\mathrm{T}} \equiv (\mathcal{N}\sigma_{\mathrm{T}})^{-1}, \tag{1.85}$$

is the (total) mean free path between interactions.

Let us now consider a particle that moves within an unbound medium. The PDF $p(s)$ of the path length s of the particle from its current position to the site of the next interaction may be obtained as follows. The probability that the particle travels a path length s without interacting is

$$\mathcal{F}(s) = \int_s^\infty p(s')\,\mathrm{d}s'. \tag{1.86}$$

The probability $p(s)\,\mathrm{d}s$ of having the next interaction when the travelled length is in the interval $(s, s+\mathrm{d}s)$ equals the product of $\mathcal{F}(s)$ (the probability of arrival at s without interacting) and $\lambda_{\mathrm{T}}^{-1}\,\mathrm{d}s$ (the probability of interacting within $\mathrm{d}s$). It then follows that

$$p(s) = \lambda_{\mathrm{T}}^{-1} \int_s^\infty p(s')\,\mathrm{d}s'. \tag{1.87}$$

The solution of this integral equation, with the boundary condition $p(\infty) = 0$, is the familiar exponential distribution

$$p(s) = \lambda_{\mathrm{T}}^{-1} \exp\left(-s/\lambda_{\mathrm{T}}\right). \tag{1.88}$$

Notice that the mean free path λ_{T} coincides with the average path length between collisions:

$$\langle s \rangle = \int_0^\infty s\,p(s)\,\mathrm{d}s = \lambda_{\mathrm{T}}. \tag{1.89}$$

The differential inverse mean free path for the interaction process A is defined as

$$\frac{\mathrm{d}^2\lambda_{\mathrm{A}}^{-1}}{\mathrm{d}W\mathrm{d}\Omega}(E; W, \theta) = \mathcal{N}\frac{\mathrm{d}^2\sigma_{\mathrm{A}}}{\mathrm{d}W\mathrm{d}\Omega}(E; W, \theta). \tag{1.90}$$

Evidently, the integral of the differential inverse mean free path gives the inverse mean free path for the process,

$$\lambda_{\mathrm{A}}^{-1} = \int \mathrm{d}W \int 2\pi \sin\theta \, \mathrm{d}\theta \, \frac{\mathrm{d}^2\lambda_{\mathrm{A}}^{-1}}{\mathrm{d}W \mathrm{d}\Omega}(E; W, \theta) = \mathcal{N}\sigma_{\mathrm{A}}. \tag{1.91}$$

In the literature, the product $\mathcal{N}\sigma_{\mathrm{A}}$ is frequently called the *macroscopic cross section*, although this name is not appropriate for a quantity that has the dimensions of inverse length. Notice that the total inverse mean free path is the sum of the inverse mean free paths of the different active interaction mechanisms,

$$\lambda_{\mathrm{T}}^{-1} = \lambda_{\mathrm{A}}^{-1} + \lambda_{\mathrm{B}}^{-1}. \tag{1.92}$$

1.4.2 Generation of random tracks

Each particle track starts off at a given position, with initial direction and energy in accordance with the characteristics of the source. The "state" of a particle immediately after an interaction (or after entering the sample or starting its trajectory) is defined by its position coordinates $\mathbf{r} = (x, y, z)$, energy E and direction cosines of the direction of flight, i.e. the components of the unit vector $\hat{\mathbf{d}} = (u, v, w)$, as seen from the laboratory reference frame. Each simulated track is thus characterized by a series of states $\mathbf{r}_n, E_n, \hat{\mathbf{d}}_n$, where $\mathbf{r}_n$ is the position of the n-th scattering event and E_n and $\hat{\mathbf{d}}_n$ are the energy and direction cosines of the direction of movement just *after* that event.

The generation of random tracks proceeds as follows. Let us assume that a track has already been simulated up to a state $\mathbf{r}_n, E_n, \hat{\mathbf{d}}_n$. The length s of the free path to the next collision, the involved scattering mechanism, the change of direction and the energy loss in this collision are random variables that are sampled from the corresponding PDFs, using the methods described in section 1.2. Hereafter, ξ stands for a random number uniformly distributed in the interval (0,1).

The length of the free flight is distributed according to the PDF given by eq. (1.88). Random values of s are generated by using the sampling formula [see eq. (1.36)]

$$s = -\lambda_{\mathrm{T}} \ln \xi. \tag{1.93}$$

The following interaction occurs at the position

$$\mathbf{r}_{n+1} = \mathbf{r}_n + s\hat{\mathbf{d}}_n. \tag{1.94}$$

The type of this interaction ("A" or "B") is selected from the point probabilities given by eq. (1.84) using the inverse transform method (section 1.2.2). The energy loss W and the polar scattering angle θ are sampled from the distribution $p_{\mathrm{A,B}}(E; W, \theta)$, eq. (1.81), by using a suitable sampling technique. The azimuthal scattering angle is generated, according to the uniform distribution in $(0, 2\pi)$, as $\phi = 2\pi\xi$.

After sampling the values of W, θ and ϕ, the energy of the particle is reduced, $E_{n+1} = E_n - W$, and the direction of movement after the interaction $\hat{\mathbf{d}}_{n+1} = (u', v', w')$

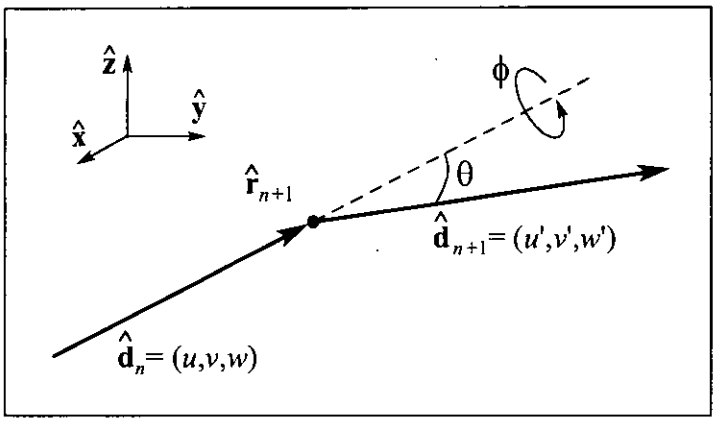


Figure 1.6: Angular deflections in single-scattering events.

is obtained by performing a rotation of $\hat{\mathbf{d}}_n = (u, v, w)$ (see fig. 1.6). The rotation matrix $R(\theta, \phi)$ is determined by the polar and azimuthal scattering angles. To explicitly obtain the direction vector $\hat{\mathbf{d}}_{n+1} = R(\theta, \phi)\hat{\mathbf{d}}_n$ after the interaction, we first note that, if the initial direction is along the z-axis, the direction after the collision is

$$\begin{pmatrix} \sin\theta\cos\phi \\ \sin\theta\sin\phi \\ \cos\theta \end{pmatrix} = R_z(\phi)R_y(\theta)\hat{\mathbf{z}}, \tag{1.95}$$

where $\hat{\mathbf{z}}$=(0,0,1) and

$$R_y(\theta) = \begin{pmatrix} \cos\theta & 0 & \sin\theta \\ 0 & 1 & 0 \\ -\sin\theta & 0 & \cos\theta \end{pmatrix} \quad \text{and} \quad R_z(\phi) = \begin{pmatrix} \cos\phi & -\sin\phi & 0 \\ \sin\phi & \cos\phi & 0 \\ 0 & 0 & 1 \end{pmatrix} \tag{1.96}$$

are rotation matrices corresponding to active rotations of angles θ and ϕ about the y- and z-axes, respectively. On the other hand, if ϑ and φ are the polar and azimuthal angles of the initial direction

$$\hat{\mathbf{d}}_n = (\sin\vartheta\cos\varphi, \sin\vartheta\sin\varphi, \cos\vartheta), \tag{1.97}$$

the rotation $R_y(-\vartheta)R_z(-\varphi)$ transforms the vector $\hat{\mathbf{d}}_n$ into $\hat{\mathbf{z}}$. It is then clear that the final direction vector $\hat{\mathbf{d}}_{n+1}$ can be obtained by performing the following sequence of rotations of the initial direction vector: 1) $R_y(-\vartheta)R_z(-\varphi)$, which transforms $\hat{\mathbf{d}}_n$ into $\hat{\mathbf{z}}$; 2) $R_z(\phi)R_y(\theta)$, which rotates $\hat{\mathbf{z}}$ according to the sampled polar and azimuthal scattering angles; and 3) $R_z(\varphi)R_y(\vartheta)$, which inverts the rotation of the first step. Hence

$$R(\theta, \phi) = R_z(\varphi)R_y(\vartheta)R_z(\phi)R_y(\theta)R_y(-\vartheta)R_z(-\varphi). \tag{1.98}$$

The final direction vector is

$$\hat{\mathbf{d}}_{n+1} = R(\theta,\phi)\hat{\mathbf{d}}_n = R_z(\varphi)R_y(\vartheta)\begin{pmatrix}\sin\theta\cos\phi\\ \sin\theta\sin\phi\\ \cos\theta\end{pmatrix} \tag{1.99}$$

and its direction cosines are

$$\begin{aligned} u' &= u\cos\theta + \frac{\sin\theta}{\sqrt{1-w^2}}\left[uw\cos\phi - v\sin\phi\right],\\ v' &= v\cos\theta + \frac{\sin\theta}{\sqrt{1-w^2}}\left[vw\cos\phi + u\sin\phi\right],\\ w' &= w\cos\theta - \sqrt{1-w^2}\sin\theta\cos\phi. \end{aligned} \tag{1.100}$$

These equations are indeterminate when $w \simeq \pm 1$, i.e. when the initial direction is nearly parallel or antiparallel to the z-axis; in this case we can simply set

$$u = \pm\sin\theta\,\cos\phi, \quad v = \pm\sin\theta\,\sin\phi, \quad w = \pm\cos\theta. \tag{1.101}$$

Moreover, eqs. (1.100) are not very stable numerically and the normalization of $\hat{\mathbf{d}}_{n+1}$ tends to drift from 1 after repeated usage. This must be remedied by periodically renormalizing $\hat{\mathbf{d}}_{n+1}$. The change of direction expressed by eqs. (1.100) and (1.101) is performed by the subroutine `DIRECT` (see the PENELOPE source listing).

The simulation of the track then proceeds by repeating these steps. A track is finished either when it leaves the material system or when the energy becomes smaller than a given energy $E_{\rm abs}$, which is the energy where particles are assumed to be effectively stopped and absorbed in the medium.

1.4.3 Particle transport as a Markov process

The foregoing concepts, definitions and simulation scheme rest on the assumption that particle transport can be modelled as a Markov process[2], i.e. "future values of a random variable (interaction event) are statistically determined by present events and depend only on the event immediately preceeding". Owing to the Markovian character of the transport, we can stop the generation of a particle history at an arbitrary state (any point of the track) and resume the simulation from this state without introducing any bias in the results.

In mixed simulations of electron/positron transport, it is necessary to limit the length s of each "free jump" so that it does not exceed a given value $s_{\rm max}$. To accomplish this, we still sample the free path length s to the next interaction from the exponential PDF

[2]The quoted definition is from the Webster's Encyclopedic Unabridged Dictionary of the English Language (Portland House, New York, 1989).

(1.88), but when $s > s_{\max}$ we only let the particle advance a distance $s_{\max}$ along the direction of motion. At the end of the truncated free jump we do nothing (i.e. the particle keeps its energy and direction of motion unaltered); however, for programming convenience, we shall say that the particle suffers a *delta interaction* (actually, a "non-interaction"). When the sampled value of s is less than $s_{\max}$, a real interaction is simulated. After the interaction (either real or delta), we sample a new free path s, move the particle a distance $s' = \min(s, s_{\max})$, etc. From the Markovian character of the transport, it is clear that the insertion of delta interactions keeps the simulation unbiased. If you do not see it so clearly, here comes a direct proof. First we note that the probability that a free jump ends with a delta interaction is

$$p_\delta = \int_{s_{\max}}^{\infty} p(s)\,\mathrm{d}s = \exp(-s_{\max}/\lambda_{\mathrm{T}}). \tag{1.102}$$

To obtain the probability $p(s)\mathrm{d}s$ of having the first real interaction at a distance in the interval $(s, s + \mathrm{d}s)$, we write $s = ns_{\max} + s'$ with $n = [s/s_{\max}]$ and, hence, $s' < s_{\max}$. The sought probability is then equal to the probability of having n successive delta interactions followed by a real interaction at a distance in $(s', s' + \mathrm{d}s)$ from the last, n-th, delta interaction,

$$p(s)\,\mathrm{d}s = p_\delta^n\ \lambda_{\mathrm{T}}^{-1} \exp(-s'/\lambda_{\mathrm{T}})\,\mathrm{d}s = \lambda_{\mathrm{T}}^{-1} \exp(-s/\lambda_{\mathrm{T}})\,\mathrm{d}s, \tag{1.103}$$

which is the correct value [cf. eq. (1.88)].

Up to this point, we have considered transport in a single homogeneous medium. In practical cases, however, the material structure where radiation is transported may consist of various regions with different compositions. We assume that the interfaces between contiguous media are sharp (i.e. there is no diffusion of chemical species across them) and passive (which amounts to neglecting e.g. surface plasmon excitation and transition radiation). In the simulation code, when a particle arrives at an interface, it is stopped there and the simulation is resumed with the interaction properties of the new medium. Obviously, this procedure is consistent with the Markovian property of the transport process.

Consider two homogeneous media, 1 and 2 (with corresponding mean free paths $\lambda_{\mathrm{T},1}$ and $\lambda_{\mathrm{T},2}$), separated by an interface, which is crossed by particles that move from the first medium to the second. The average path length between the last real interaction in medium 1 and the first real interaction in medium 2 is $\lambda_{\mathrm{T},1} + \lambda_{\mathrm{T},2}$, as can be easily verified by simulation. This result seemed paradoxical to some authors and induced confusion in the past. In fact, there is nothing odd here as you may easily verify (again by simulation) as follows. Assume particles being transported within a single homogeneous medium with an imaginary plane that acts as a "virtual" interface, splitting the medium into two halves. In the simulation, the particles do not see this interface, i.e. they do not stop when crossing. Every time a particle crosses the plane, we score the length s_{plane} of the track segment between the two real interactions immediately before and after the crossing. It is found that the average value of s_{plane} is $2\lambda_{\mathrm{T}}$, in spite of the fact that the free path length between consecutive collisions was sampled from an exponential PDF

with the mean free path λ_T [yes, the scored values s_{plane} were generated from this PDF!]. The explanation of this result is that, as a consequence of the Markovian character, the average path length from the plane (an arbitrary *fixed* point in the track) back to the last collision (or up to the next collision) is λ_T.

1.5 Statistical averages and uncertainties

For the sake of being more specific, let us consider the simulation of a high-energy electron beam impinging on the surface of a semi-infinite water phantom. Each primary electron originates a shower of electrons and photons, which are individually tracked down to the corresponding absorption energy. Any quantity of interest Q is evaluated as the average score of a large number N of simulated random showers. Formally, Q can be expressed as an integral of the form (1.64),

$$Q = \int q\, p(q)\, \mathrm{d}q, \tag{1.104}$$

where the PDF $p(q)$ is usually unknown. The simulation of individual showers provides a practical method to sample q from the "natural" PDF $p(q)$: from each generated shower we get a random value q_i distributed according to $p(q)$. The only difference to the case of Monte Carlo integration considered above is that now the PDF $p(q)$ describes a cascade of random interaction events, each with its characteristic PDF. The Monte Carlo estimate of Q is

$$\overline{Q} = \frac{1}{N}\sum_{i=1}^{N} q_i. \tag{1.105}$$

Thus, for instance, the average energy E_{dep} deposited within the water phantom per incident electron is obtained as

$$E_{\text{dep}} = \frac{1}{N}\sum_{i=1}^{N} e_i, \tag{1.106}$$

where e_i is the energy deposited by *all* the particles of the i-th shower. The statistical uncertainty (standard deviation) of the Monte Carlo estimate [eq. (1.72)] is

$$\sigma_Q = \sqrt{\frac{\mathrm{var}(q)}{N}} = \sqrt{\frac{1}{N}\left[\frac{1}{N}\sum_{i=1}^{N} q_i^2 - \overline{Q}^2\right]}. \tag{1.107}$$

As mentioned above, we shall usually express the simulation result in the form $\overline{Q} \pm 3\sigma_Q$, so that the interval $(\overline{Q} - 3\sigma_Q,\ \overline{Q} - 3\sigma_Q)$ contains the true value Q with 99.7% probability. Notice that to evaluate the standard deviation (1.107) we must score the squared contributions q_i^2. In certain cases, the contributions q_i can only take the values 0 and 1, and the standard error can be determined without scoring the squares,

$$\sigma_Q = \sqrt{\frac{1}{N}\overline{Q}(1-\overline{Q})}. \tag{1.108}$$

Simulation/scoring can also be used to compute continuous distributions. The simplest method is to "discretize" the distributions, by treating them as histograms, and to determine the "heights" of the different bars. To make the arguments clear, let us consider the depth-dose distribution $D(z)$, defined as the average energy deposited per unit depth and per incident electron within the water phantom. $D(z)\mathrm{d}z$ is the average energy deposited at depths between z and $z+\mathrm{d}z$ per incident electron, and the integral of $D(z)$ from 0 to ∞ is the average deposited energy E_{dep} (again, per incident electron). Since part of the energy is reflected back from the water phantom (through backscattered radiation), E_{dep} is less than the kinetic energy E_{inc} of the incident electrons. We are interested in determining $D(z)$ in a limited depth interval, say from $z = 0$ to $z = z_{\mathrm{max}}$. The calculation proceeds as follows. First of all, we have to select a partition of the interval $(0, z_{\mathrm{max}})$ into M different depth bins (z_{k-1}, z_k), with $0 = z_0 < z_1 < \ldots < z_M = z_{\mathrm{max}}$. Let $e_{ij,k}$ denote the amount of energy deposited into the k-th bin by the j-th particle of the i-th shower (each incident electron may produce multiple secondary particles). The average energy deposited into the k-th bin (per incident electron) is obtained as

$$E_k = \frac{1}{N}\sum_{i=1}^{N} e_{i,k} \qquad \text{with} \qquad e_{i,k} \equiv \sum_j e_{ij,k}, \tag{1.109}$$

and is affected by a statistical uncertainty

$$\sigma_{Ek} = \sqrt{\frac{1}{N}\left[\frac{1}{N}\sum_{i=1}^{N} e_{i,k}^2 - E_k^2\right]}. \tag{1.110}$$

The Monte Carlo depth-dose distribution $D_{\mathrm{MC}}(z)$ is a stepwise constant function,

$$D_{\mathrm{MC}}(z) = D_k \pm 3\sigma_{Dk} \qquad \text{for } z_{k-1} < z < z_k \tag{1.111}$$

with

$$D_k \equiv \frac{1}{z_k - z_{k-1}} E_k, \qquad \sigma_{Dk} \equiv \frac{1}{z_k - z_{k-1}} \sigma_{Ek}. \tag{1.112}$$

Notice that the bin average and standard deviation have to be divided by the bin width to obtain the final Monte Carlo distribution. Defined in this way, $D_{\mathrm{MC}}(z)$ is an unbiased estimator of the *average* dose in each bin. The limitation here is that we are approximating the continuous distribution $D(z)$ as a histogram with finite bar widths. In principle, we could obtain a closer approximation by using narrower bins. However, care has to be taken in selecting the bin widths since statistical uncertainties may completely hide the information in narrow bins.

A few words regarding programming details are in order. To evaluate the average deposited energy and its standard deviation for each bin, eqs. (1.109) and (1.110), we must score the shower contributions $e_{i,k}$ and their squares $e_{i,k}^2$. There are cases in which a senseless literal application of this recipe may take a large fraction of the simulation time. Consider, for instance, the simulation of the 3D dose distribution in the phantom, which may involve several thousand volume bins. For each bin, the energies $e_{ij,k}$ deposited by the individual particles of a shower must be accumulated in a partial counter to obtain

the shower contribution $e_{i,k}$ and, after completion of the whole shower, the value $e_{i,k}$ and its square must be added to the accumulated counters. As only a small fraction of the bins receive energy from a single shower, it is not practical to treat all bin counters on an equal footing. The fastest method is to transfer partial scores to the accumulated counters only when the partial counter is going to receive a contribution from a new shower. This can be easily implemented in a computer program as follows. For each quantity of interest, say Q, we define three real counters, `Q`, `Q2` and `QP`, and an integer label `LQ`; all these quantities are initially set to zero. The partial scores q_{ij} of the particles of a shower are accumulated in the partial counter `QP`, whereas the global shower contribution q_i and its square are accumulated in `Q` and `Q2`, respectively. Each shower is assigned a label, for instance its order number i, which is stored in `LQ` the first time that the shower contributes to `QP`. In the course of the simulation, the value of `QP` is transferred to the global counters `Q` and `Q2` only when it is necessary to store a contribution q_{ij} from a new shower. Explicitly, the FORTRAN code for scoring Q is

```
IF(i.NE.LQ) THEN
  Q=Q+QP
  Q2=Q2+QP**2
  QP=q_ij
  LQ=i
ELSE
  QP=QP+q_ij
ENDIF
```

At the end of the simulation, the residual contents of `QP` must be transferred to the global counters.

For some quantities (e.g. the mean number of scattering events per track, the depth-dose function, ...) almost all the simulated tracks contribute to the score and the inherent statistical uncertainties of the simulation results are comparatively small. Other quantities (e.g. angle and energy distributions of the particles transmitted through a thick foil) have considerable statistical uncertainties (i.e. large variances) because only a small fraction of the simulated tracks contribute to the partial scores.

1.6 Variance reduction

In principle, the statistical error of a quantity may be somewhat reduced (without increasing the computer simulation time) by using variance-reduction techniques. Unfortunately, these optimization techniques are extremely problem-dependent, and general recipes to minimize the variance cannot be given. On the other hand, the importance of variance reduction should not be overvalued. In many cases, analogue[3] simulation does the work in a reasonable time. Spending manhours by complicating the program, to get a modest reduction in computing time may not be a good investment. It is

[3]We use the term "analogue" to refer to detailed, condensed or mixed simulations that do not incorporate variance-reduction procedures.

also important to realize that an efficient variance-reduction method usually lowers the statistical error of a given quantity Q at the expense of increasing the uncertainties of other quantities. Thus, variance-reduction techniques are not recommended when a global description of the transport process is sought. Here we give a brief description of those techniques which, with a modest programming effort, can be useful in improving the solution of some ill-conditioned problems. For the sake of generality, we consider that secondary particles can be generated in the interactions with the medium. A nice, and practically oriented, review of variance-reduction methods in radiation transport has been given by Bielajew and Rogers (1988).

1.6.1 Interaction forcing

Sometimes, a high variance results from an extremely low interaction probability. Consider, for instance, the simulation of the energy spectrum of bremsstrahlung photons emitted by medium energy ($\sim$ 100 keV) electrons in a thin foil of a certain material. As radiative events are much less probable than elastic and inelastic scattering, the uncertainty of the simulated photon spectrum will be relatively large. In such cases, an efficient variance-reduction method is to artificially increase the interaction probability of the process A of interest. Our practical implementation of interaction forcing consists of replacing the mean free path $\lambda_{\rm A}$ of the real process by a shorter one, $\lambda_{\rm A,f}$, i.e. we force A interactions to occur more frequently than for the real process. We consider that the PDF for the energy loss, the angular deflections (and the directions of emitted secondary particles, if any) in the forced interactions is the same as for the real interactions. To sample the length of the free jump to the next interaction, we use the exponential distribution with the reduced mean free path $\lambda_{\rm A,f}$. This is equivalent to increasing the interaction probability per unit path length of the process A by a factor

$$\mathcal{F} = \frac{\lambda_{\rm A}}{\lambda_{\rm A,f}} > 1. \tag{1.113}$$

To keep the simulation unbiased, we must correct for the introduced distortion as follows:

(i) A weight $w_{\rm p}^{(1)} = 1$ is associated with each primary particle. Secondary particles produced in forced interactions have an associated weight $w_{\rm p}^{(2)} = w_{\rm p}^{(1)}/\mathcal{F}$; the weights of successive generations of forced secondaries are $w_{\rm p}^{(k)} = w_{\rm p}^{(k-1)}/\mathcal{F}$. Secondary particles generated in non-forced interactions (i.e. of types other than A) are given a weight equal to that of their parent particle.

(ii) A weight $w_E^{(k)} = w_{\rm p}^{(k)}/\mathcal{F}$ is given to the deposited energy (and to any other alteration of the medium such as e.g. charge deposition) that results from forced interactions of a particle with weight $w_{\rm p}^{(k)}$. For non-forced interactions $w_E^{(k)} = w_{\rm p}^{(k)}$.

(iii) Forced interactions are simulated to determine the energy loss and possible emission of secondary radiation, but the state variables of the interacting particle are altered only with probability $1/\mathcal{F}$. That is, the energy E and direction of movement $\hat{\mathbf{d}}$ of the projectile are varied only when the value ξ of a random number falls below $1/\mathcal{F}$, otherwise E and $\hat{\mathbf{d}}$ are kept unchanged.

Of course, interaction forcing should be applied only to interactions that are dynamically allowed, i.e. for particles with energy above the corresponding "reaction" threshold.

Let w_{i1} and q_{i1} denote the weight and the contribution to the score of the i-th primary, and let w_{ij} and q_{ij} $(j > 1)$ represent the weights and contributions of the j-th secondary particles generated by the i-th primary. The Monte Carlo estimate of Q obtained from the N simulated histories is

$$\overline{Q} = \frac{1}{N} \sum_{i,j} w_{ij} q_{ij}. \tag{1.114}$$

Evidently, the estimates $\overline{Q}$ obtained with interaction forcing and from an analogue simulation are equal (in the statistical sense, i.e. in the limit $N \to \infty$, their difference tends to zero). The standard deviation is given by

$$\sigma_Q = \sqrt{\frac{1}{N}\left[\frac{1}{N}\sum_i \left(\sum_j w_{ij} q_{ij}\right)^2 - \overline{Q}^2\right]}. \tag{1.115}$$

Quantities directly related to the forced interactions will have a reduced statistical error, due to the increase in number of these interactions. However, for a given simulation time, other quantities may exhibit standard deviations larger than those of the analogue simulation, because of the time spent in simulating the forced interactions.

1.6.2 Splitting and Russian roulette

These two techniques, which are normally used in conjunction, are effective in problems where interest is focused on a localized spatial region. Typical examples are the calculation of dose functions in deep regions of irradiated objects and, in the case of collimated radiation beams, the evaluation of radial doses far from the beam axis. The basic idea of splitting and Russian roulette methods is to favour the flux of radiation towards the region of interest and inhibit the radiation that leaves that region. These techniques are also useful in other problems where only a partial description of the transport process is required. The "region of interest" may then be a limited volume in the space of state variables $(\mathbf{r}, E, \hat{\mathbf{d}})$. Thus, in studies of radiation backscattering, the region of interest may be selected as the spatial region of the sample close to the irradiated surface *and* the set of particle directions that point towards this surface.

As in the case of interaction forcing, variance reduction is accomplished by modifying the weights of the particles. It is assumed that primary particles start moving with unit weight and each secondary particle produced by a primary one is assigned an initial weight equal to that of the primary. Splitting consists of transforming a particle, with weight w_0 and in a certain state, into a number $\mathcal{S} > 1$ of identical particles with weights $w = w_0/\mathcal{S}$ in the same initial state. Splitting should be applied when the particle "approaches" the region of interest. The Russian roulette technique is, in a way, the reverse process: when a particle tends to move away from the region of interest it is

"killed" with a certain probability, $\mathcal{K} < 1$, and, if it survives, its weight is increased by a factor $1/(1-\mathcal{K})$. Here, killing means that the particle is just discarded (and does not contribute to the scores anymore). Evidently, splitting and killing leave the simulation unbiased. The mean and standard deviation of the calculated quantities are given by eqs. (1.114) and (1.115). The effectiveness of these methods relies on the adopted values of the parameters $\mathcal{S}$ and $\mathcal{K}$, and on the strategy used to decide when splitting and killing are to be applied. These details can only be dictated by the user's experience.

1.6.3 Other methods

Very frequently, an effective "reduction of variance" may be obtained by simply avoiding unnecessary calculations. This is usually true for simulation codes that incorporate "general-purpose" geometry packages. In the case of simple (e.g. planar, spherical, cylindrical) geometries the program may be substantially simplified and this may speed up the simulation appreciably. In general, the clever use of possible symmetries of the problem under consideration may lead to spectacular variance reductions. As a last example, we can quote the so-called "range rejection" method, which simply consists of absorbing a particle when it (and its possible secondaries) cannot leave (or reach) the regions of interest. Range rejection is useful e.g. when computing the total energy deposition of electrons or positrons in a given spatial region. When the residual range of a particle is less than the distance to the nearest limiting surface of the region of interest, the particle will deposit all its energy inside or outside the considered region (depending of its current position) and the simulation of the track can be stopped. Range rejection is not adequate for photon transport simulation, since the concept of photon range is not well defined (or, to be more precise, photon path length fluctuations are very large).

Chapter 2

Photon interactions

In this chapter, we consider the interactions of unpolarized photons of energy E with atoms of atomic number Z. We limit our considerations to the energy range from 100 eV up to 1 GeV, where the dominant interaction processes are coherent (Rayleigh) scattering, incoherent (Compton) scattering, the photoelectric effect and electron-positron pair production. Other interactions, such as photonuclear absorption, occur with much smaller probability and can be disregarded for most practical purposes (see e.g. Hubbell et al., 1980).

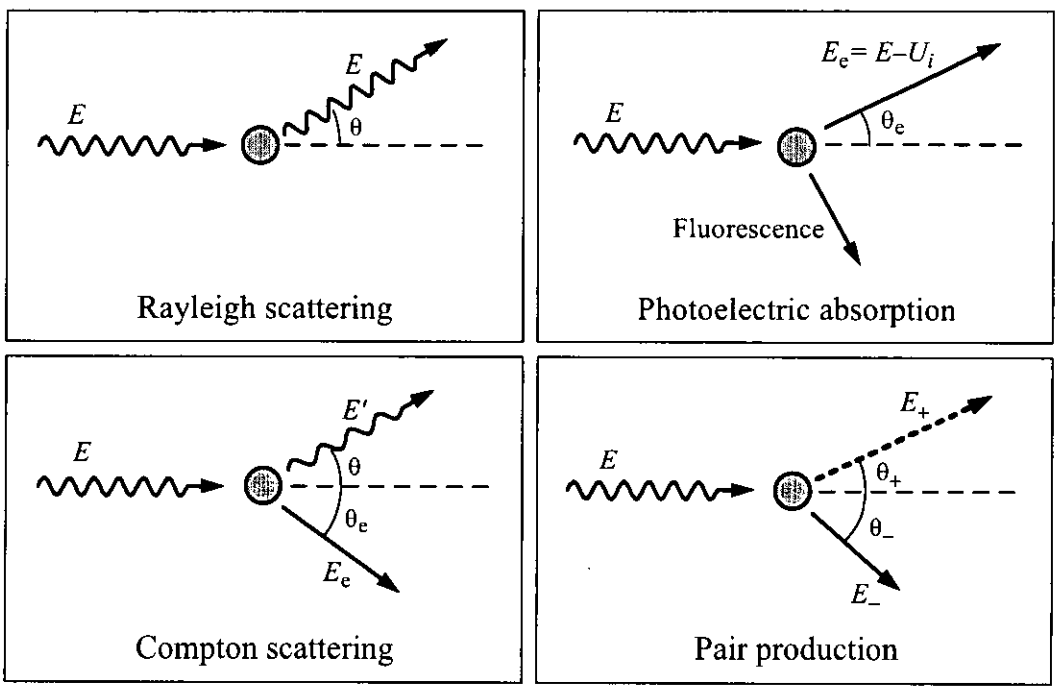


Figure 2.1: Basic interactions of photons with matter.

As long as the response of an atom is not appreciably distorted by molecular binding, the single-atom theory can be extended to molecules by using the additivity approximation, i.e. the molecular cross section for a process is approximated by the sum of the atomic cross sections of all the atoms in the molecule. The additivity approximation can also be applied to dense media whenever interference effects between waves scattered by different centres (which, for instance, give rise to Bragg diffraction in crystals) are small. We assume that these conditions are always satisfied.

The ability of Monte Carlo simulation methods to describe photon transport in complex geometries has been established from research during the last five decades (Hayward and Hubbell, 1954; Zerby, 1963; Berger and Seltzer, 1972; Chan and Doi, 1983; Ljungberg and Strand, 1989). The most accurate DCSs available are given in numerical form and, therefore, advanced Monte Carlo codes make use of extensive databases. To reduce the amount of required numerical information, in PENELOPE we use a combination of analytical DCSs and numerical tables. The adopted DCSs are defined by simple, but physically sound analytical forms. The corresponding total cross sections are obtained by a single numerical quadrature that is performed very quickly, even on a personal computer, using the SUMGA external function described in appendix B. Moreover, the random sampling from these DCSs can be done analytically and, hence, exactly. Only coherent scattering requires a simple preparatory numerical step.

It may be argued that using analytical *approximate* DCSs, instead of more accurate tabulated DCSs implies a certain loss of accuracy. To minimize this loss, PENELOPE renormalizes the analytical DCSs so as to reproduce partial attenuation coefficients that are read from the input material data file. As a consequence, the free path between events and the kind of interaction are sampled using total cross sections that are nominally exact; approximations are introduced only in the description of individual interaction events.

In the following, κ stands for the photon energy in units of the electron rest energy, i.e.

$$\kappa \equiv \frac{E}{m_e c^2}. \tag{2.1}$$

2.1 Coherent (Rayleigh) scattering

Coherent or Rayleigh scattering is the process by which photons are scattered by bound atomic electrons without excitation of the target atom, i.e. the energies of the incident and scattered photons are the same. The scattering is qualified as "coherent" because it arises from the interference between secondary electromagnetic waves coming from different parts of the atomic charge distribution.

The atomic DCS per unit solid angle for coherent scattering is given approximately by (see e.g. Born, 1969)

$$\frac{d\sigma_{Ra}}{d\Omega} = \frac{d\sigma_T}{d\Omega} [F(q, Z)]^2 , \tag{2.2}$$

where

$$\frac{\mathrm{d}\sigma_{\mathrm{T}}(\theta)}{\mathrm{d}\Omega} = r_{\mathrm{e}}^2 \frac{1+\cos^2\theta}{2} \tag{2.3}$$

is the classical Thompson DCS for scattering by a free electron at rest, θ is the polar scattering angle (see fig. 2.1) and $F(q,Z)$ is the atomic form factor. The quantity r_{e} is the classical electron radius and q is the magnitude of the momentum transfer given by

$$q = 2(E/c)\sin(\theta/2) = (E/c)\left[2(1-\cos\theta)\right]^{1/2}. \tag{2.4}$$

In the literature on x-ray crystallography, the dimensionless variable

$$x = \frac{q\ 10^{-8}\mathrm{cm}}{4\pi\hbar} = 20.6074\ \frac{q}{\mathrm{m_e}c} \tag{2.5}$$

is normally used instead of q.

The atomic form factor can be expressed as the Fourier transform of the atomic electron density $\rho(\mathbf{r})$ which, for a spherically symmetrical atom, simplifies to

$$F(q,Z) = 4\pi \int_0^\infty \rho(r) \frac{\sin(qr/\hbar)}{qr/\hbar} r^2\,\mathrm{d}r. \tag{2.6}$$

$F(q,Z)$ is a monotonically decreasing function of q that varies from $F(0,Z) = Z$ to $F(\infty,Z) = 0$. The most accurate form factors are those obtained from Hartree-Fock or configuration-interaction atomic-structure calculations; here we adopt the non-relativistic atomic form factors tabulated by Hubbell et al. (1975). Although relativistic form factors are available (Doyle and Turner, 1968), Hubbell has pointed out that the non-relativistic form factors yield results in closer agreement with experiment (Cullen et al., 1997).

In the calculations, we use the following analytical approximation

$$F(q,Z) = \begin{cases} f(x,Z) \equiv Z\ \dfrac{1+a_1x^2+a_2x^3+a_3x^4}{\left(1+a_4x^2+a_5x^4\right)^2}, \\ \max\left\{f(x,Z), F_{\mathrm{K}}(q,Z)\right\} \quad \text{if } Z>10 \text{ and } f(x,Z)<2, \end{cases} \tag{2.7}$$

where

$$F_{\mathrm{K}}(q,Z) \equiv \frac{\sin(2b\arctan Q)}{bQ\left(1+Q^2\right)^b}, \tag{2.8}$$

with

$$Q = \frac{q}{2\mathrm{m_e}ca}, \qquad b = \sqrt{1-a^2}, \qquad a \equiv \alpha(Z-5/16), \tag{2.9}$$

where α is the fine-structure constant. The function $F_{\mathrm{K}}(q,Z)$ is the contribution to the atomic form factor due to the two K-shell electrons (see e.g. Baró et al., 1994a). The parameters of expression $f(x,Z)$ for $Z = 1$ to 92, which have been determined by Baró et al. (1994a) by numerically fitting the atomic form factors tabulated by Hubbell et al. (1975), are included in the block data subprogram `PENDAT`. The average relative

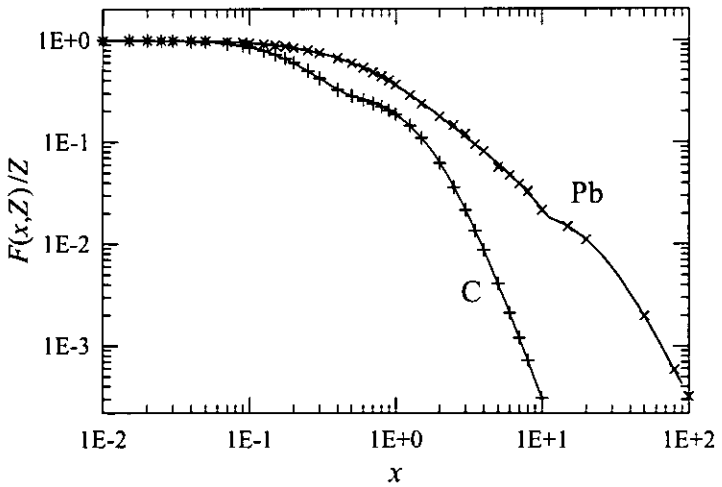


Figure 2.2: Atomic form factors for carbon and lead. Crosses are values from the tables of Hubbell et al. (1975), continuous curves represent the analytical approximation given by eq. (2.7).

difference between the analytical and tabulated form factors is less than 0.5% (see fig. 2.2).

The total coherent scattering cross section per atom is

$$\sigma_{\mathrm{Ra}} = \int \frac{\mathrm{d}\sigma_{\mathrm{Ra}}}{\mathrm{d}\Omega}\,\mathrm{d}\Omega = \pi r_{\mathrm{e}}^2 \int_{-1}^{1} \left(1+\cos^2\theta\right) [F(q,Z)]^2\,\mathrm{d}(\cos\theta). \tag{2.10}$$

Introducing q, eq. (2.4), as a new integration variable, the asymptotic behaviour of the total cross section for small and large photon energies is made clear. For low photon energies, the form factor in the integrand does not depart appreciably from the value $F(0,Z) = Z$, i.e. coherent scattering reduces to pure Thompson scattering. Consequently, we have

$$\sigma_{\mathrm{Ra}} \simeq \frac{8}{3}\pi r_{\mathrm{e}}^2\, Z^2. \tag{2.11}$$

In the high-energy limit, we get

$$\sigma_{\mathrm{Ra}} \propto E^{-2}. \tag{2.12}$$

In practice, this limiting behaviour is attained for energies of the order of $Z/2$ MeV.

Strictly speaking, expression (2.2) is adequate only for photons with energy well above the K absorption edge. The low-energy behaviour given by eq. (2.11) is substantially altered when anomalous scattering factors are introduced (see e.g. Cullen et al., 1989; Kane et al., 1986). These factors lead to a general decrease of the coherent scattering cross section near the absorption edges and at low energies. Nevertheless, at the energies where anomalous scattering effects become significant, coherent scattering is much less probable than photoelectric absorption (see fig. 2.10 below), and the approximation given by eq. (2.2) is usually sufficient for simulation purposes.

2.1.1 Simulation of coherent scattering events

The PDF of the angular deflection, $\cos\theta$, can be written as [see eqs. (2.2) and (2.3); normalization is irrelevant here]

$$p_{\mathrm{Ra}}(\cos\theta) = \frac{1+\cos^2\theta}{2}\left[F(x,Z)\right]^2, \tag{2.13}$$

where x, which is defined by eqs. (2.4) and (2.5), can take values in the interval from 0 to

$$x_{\max} = 20.6074 \times 2\kappa. \tag{2.14}$$

This PDF can be factorized in the form

$$p_{\mathrm{Ra}}(\cos\theta) = g(\cos\theta)\pi(x^2) \tag{2.15}$$

with

$$g(\cos\theta) \equiv \frac{1+\cos^2\theta}{2} \qquad \text{and} \qquad \pi(x^2) \equiv \left[F(x,Z)\right]^2. \tag{2.16}$$

Notice that, for a compound, $[F(x,Z)]^2$ has to be replaced by the sum of squared form factors of the atoms in the molecule.

The function $\pi(x^2)$ can be considered as the (unnormalized) PDF of the variable x^2. Random values of x^2 distributed according to this PDF can be generated by the inverse transform method (section 1.2.2), i.e. from the sampling equation

$$\int_0^{x^2} \pi(x'^2)\,\mathrm{d}x'^2 = \xi \int_0^{x^2_{\max}} \pi(x'^2)\,\mathrm{d}x'^2. \tag{2.17}$$

It is convenient to introduce the function

$$\Pi(x^2) = \int_0^{x^2} \pi(x'^2)\,\mathrm{d}x'^2, \tag{2.18}$$

which increases monotonically with x^2 and saturates for high x^2-values to a constant finite value. Then, the sampling equation (2.17) can be written in the form

$$\Pi(x^2) = \xi\Pi(x^2_{\max}), \tag{2.19}$$

which is easy to solve numerically. To this end, we only need to have a table of values of the function $\Pi(x^2)$ stored in memory. For a given photon energy, $\Pi(x^2_{\max})$ can be evaluated by interpolation in this table. Linear log-log interpolation (extrapolation) in a table with about 240 points logarithmically distributed in the interval $(10^{-4},10^6)$ yields results which are accurate to within 0.01% (notice that in the interval from 0 to 10^{-4}, $F(x,Z) \simeq Z$ and, hence, $\Pi(x^2)$ is proportional to x^2, i.e. extrapolation for $x^2 < 10^{-4}$ is exact). The value

$$x^2 = \Pi^{-1}\left(\xi\Pi(x^2_{\max})\right) \tag{2.20}$$

can then be obtained by inverse linear interpolation (or extrapolation) with a binary search.

The angular deflection $\cos\theta$ can now be sampled by the rejection method (section 1.2.4), since the function $g(\cos\theta)$ is a valid rejection function (i.e. it is positive and less than or equal to unity). The algorithm for sampling $\cos\theta$ proceeds as follows:

(i) Compute $\Pi(x^2_{\max})$.

(ii) Generate a random number ξ and determine x^2 using eq. (2.20). Set

$$\cos\theta = 1 - \frac{1}{2}\frac{x^2}{(20.6074\kappa)^2}. \tag{2.21}$$

(iii) Generate a new random number ξ.

(iv) If $\xi > g(\cos\theta)$, go to step (ii).

(v) Deliver $\cos\theta$.

Although numerical interpolation is necessary, it is performed on a single function that is independent of the photon energy and the errors introduced are negligible. It is worth noting that the sampling algorithm is essentially independent of the adopted form factor, and directly applicable to molecules. The advantage of using the analytical form factor, eq. (2.7), instead of a numerical database is that $\Pi(x^2)$ can be easily calculated to the desired accuracy, using the `SUMGA` integration function (appendix B).

The efficiency of the sampling method (i.e. the fraction of generated values of $\cos\theta$ that is accepted) increases with photon energy. At low energies, it equals 2/3 (exactly) for all elements. For $E = 100$ keV, the efficiencies for hydrogen and uranium are 100% and 86%, respectively.

2.2 Photoelectric effect

In the photoelectric effect, a photon of energy E is absorbed by the target atom, which makes a transition to an excited state. The photon beams found in radiation transport studies have relatively low photon densities and, as a consequence, only single-photon absorption is observed[1]. To represent the atomic states, we can adopt an independent-electron model, such as the Dirac-Hartree-Fock-Slater self-consistent model (see e.g. Pratt et al., 1973), in which each electron occupies a single-particle orbital, with well-defined ionization energy. The set of orbitals with the same principal and total angular momentum quantum numbers and the same parity constitute a shell. Each shell i can accommodate a finite number of electrons, with characteristic ionization energy U_i. Notice that the shell ionization energies are positive, the quantity $-U_i$ represents the "binding" energy of each individual electron. Fig. 2.3 (left diagram) shows the various notations used to designate the innermost atomic electron shells (i.e. those with the

[1] In intense low-energy photon beams, such as those from high-power lasers, simultaneous absorption of several photons is possible.

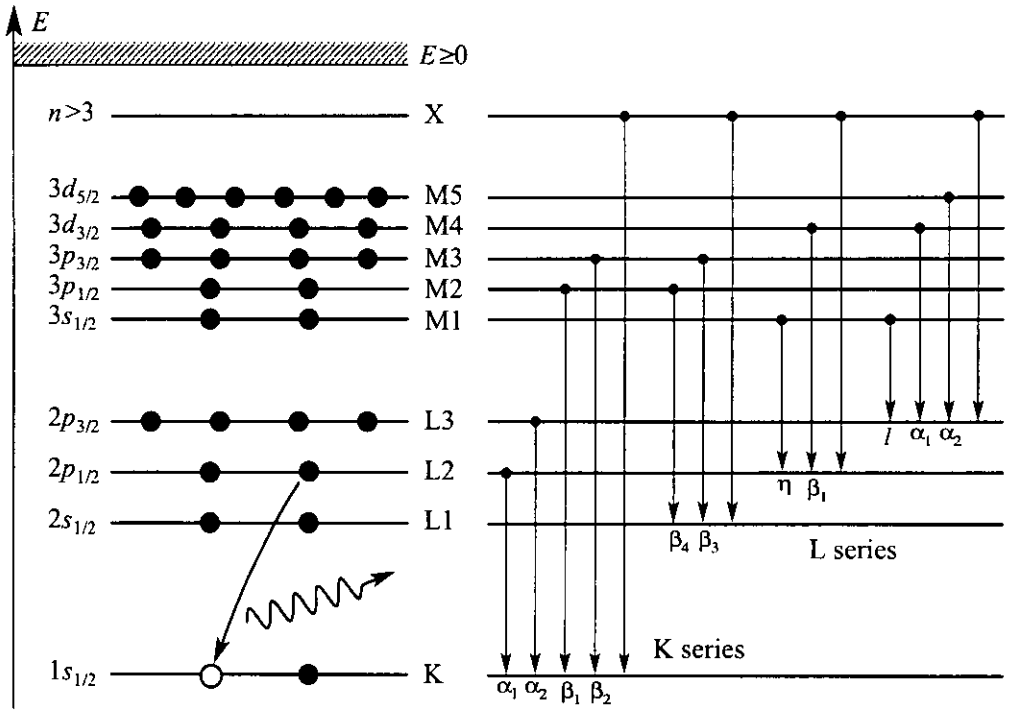


Figure 2.3: Various notations for inner atomic electron shells (left) and allowed radiative transitions (right) to these shells. Transitions different from the ones indicated in the diagram (e.g. K-M4) are also possible, but their transition probabilities are extremely small.

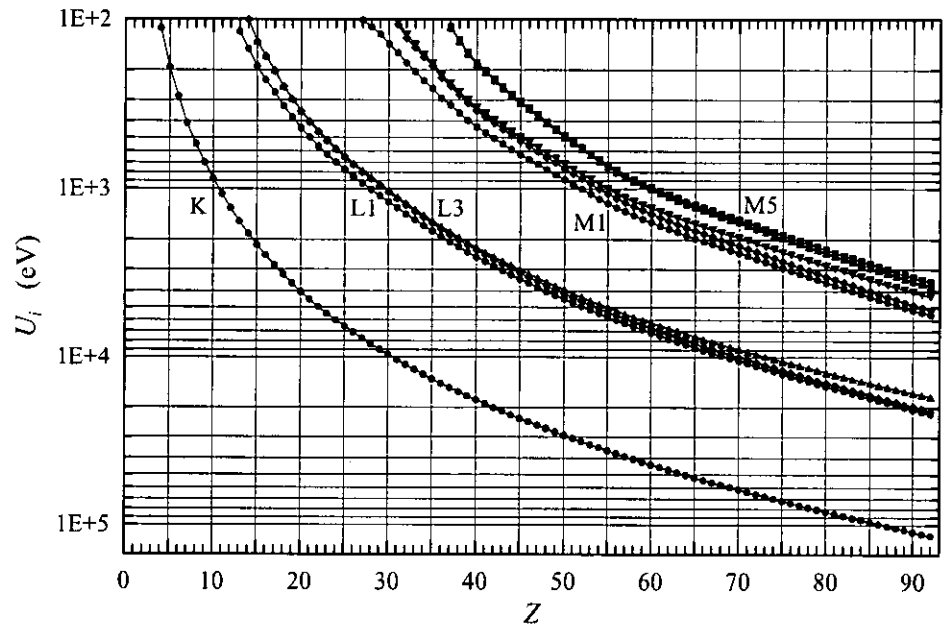


Figure 2.4: Ionization energies of the innermost shells of free atoms, as given by Lederer and Shirley (1978).

largest ionization energies) as well as their ordering in energy and allowed occupancies. In our simulations, we use the experimental ionization energies given by Lederer and Shirley (1978), which pertain to free, neutral atoms. Ionization energies of K-, L- and M-shells are displayed in fig. 2.4.

Considering the interaction with the photon field as a first-order perturbation (which is appropriate for fields with low photon densities) it follows that only one-electron transitions are allowed. That is, in the photoelectric effect, the photon is absorbed by an individual electron in the "active" shell i, which leaves the parent atom with kinetic energy $E_e = E - U_i$. Evidently, photoionization of a given shell is only possible when the photon energy exceeds the corresponding ionization energy; this gives rise to the characteristic absorption edges in the photoelectric cross section (see fig. 2.5).

The photoelectric cross sections used in PENELOPE are obtained by interpolation in a numerical table that was extracted from the LLNL Evaluated Photon Data Library (EPDL; Cullen et al., 1997). This library contains photoelectric cross sections for all shells of the elements $Z = 1 - 100$ and photon energies from 1 eV to 1000 GeV, derived from Scofield's theoretical calculations of shell cross sections (Saloman et al., 1988) and Hubbell's total cross sections (Hubbell et al., 1980; Berger and Hubbell, 1987). The PENELOPE database for photoelectric absorption (a subset of the EPDL) consists of tables of the total atomic cross section $\sigma_{\rm ph}(E)$ and the cross sections for the K- and L-shells, $\sigma_{{\rm ph},i}(E)$ (i = K, L1, L2 and L3) for the elements Z = 1–92, which span the energy range from 100 eV to 1000 GeV. These tables are estimated to be accurate to within a few percent for photon energies above 1 keV (Cullen et al., 1997). At lower energies, uncertainties in the data are much larger: 10–20% for 0.5 keV $< E <$ 1 keV and 100–200% for 0.1 keV $< E <$ 0.5 keV. Notice that the cross sections in the EPDL are based on free-atom theoretical calculations and, therefore, near-edge absorption structures produced by molecular or crystalline ordering (e.g. extended x-ray absorption fine-structure) are ignored.

For compound materials (and also for mixtures) the molecular cross section $\sigma_{\rm ph}(E)$ is evaluated by means of the additivity approximation, that is, as the sum of the atomic cross sections of the elements involved. In the energy range between successive absorption edges, the photoelectric cross section is a continuous function of the photon energy (see fig. 2.5). In PENELOPE, the molecular cross section is defined by means of a table of numerical values $\sigma_{\rm ph}(E_i)$ for a logarithmic grid of energies E_i, which is stored in memory. Photon mean free paths are determined by linear log-log interpolation in this table. Knowledge of the atomic cross sections is needed, only when a photoabsorption event has effectively occurred, to select the element that has been ionized (whose probability is proportional to the atomic cross section).

2.2.1 Simulation of photoelectron emission

Let us consider that a photon with energy E is absorbed by an atom of the element Z. The "active" shell i that is ionized is considered as a discrete random variable with

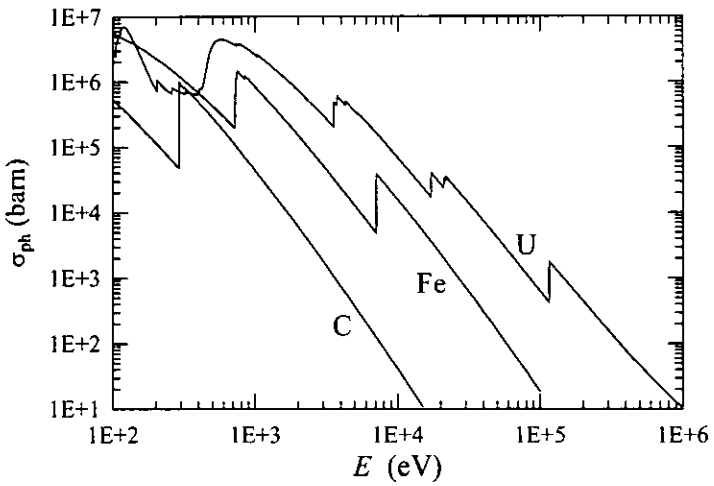


Figure 2.5: Atomic photoelectric cross sections for carbon, iron and uranium as functions of the photon energy E.

PDF

$$p_i = \sigma_{\mathrm{ph},i}(Z,E)/\sigma_{\mathrm{ph}}(Z,E), \tag{2.22}$$

where $\sigma_{\mathrm{ph},i}(Z,E)$ is the cross section for ionization of shell i and $\sigma_{\mathrm{ph}}(Z,E)$ is the total photoelectric cross section of the atom. PENELOPE incorporates a detailed description of photoabsorption in K- and L-shells (including the subsequent atomic relaxation). The ionization probabilities of these inner shells are determined from the corresponding partial cross sections. The probability of ionization in an outer shell is obtained as

$$p_{\mathrm{outer}} = 1 - (p_{\mathrm{K}} + p_{\mathrm{L1}} + p_{\mathrm{L2}} + p_{\mathrm{L3}}). \tag{2.23}$$

When the ionization occurs in an inner K- or L-shell, the initial energy of the photoelectron is set equal to $E_{\mathrm{e}} = E - U_i$; the residual atom, with a vacancy in the shell, subsequently relaxes to its ground state by emitting x-rays and Auger electrons. If the ionization occurs in an outer shell, we assume that the photoelectron leaves the target atom with kinetic energy equal to the energy deposited by the photon, $E_{\mathrm{e}} = E$, and we disregard the emission of subsidiary fluorescent radiation (see section 2.6).

Initial direction of photoelectrons

The direction of emission of the photoelectron, relative to that of the absorbed photon, is defined by the polar and azimuthal angles θ_{e} (fig. 2.1) and ϕ_{e}. We consider that the incident photon is not polarized and, hence, the angular distribution of photoelectrons is independent of ϕ_{e}, which is uniformly distributed in the interval $(0, 2\pi)$. The polar angle θ_{e} is sampled from the K-shell cross section derived by Sauter (1931) using K-shell

hydrogenic electron wave functions. The Sauter DCS (per electron) can be written as

$$\frac{\mathrm{d}\sigma_{\mathrm{ph}}}{\mathrm{d}\Omega_{\mathrm{e}}} = \alpha^4 r_{\mathrm{e}}^2 \left(\frac{Z}{\kappa}\right)^5 \frac{\beta^3}{\gamma} \frac{\sin^2\theta_{\mathrm{e}}}{(1-\beta\cos\theta_{\mathrm{e}})^4} \left[1 + \frac{1}{2}\gamma(\gamma-1)(\gamma-2)(1-\beta\cos\theta_{\mathrm{e}})\right], \quad (2.24)$$

where α is the fine-structure constant, r_{e} is the classical electron radius, and

$$\gamma = 1 + E_{\mathrm{e}}/(\mathrm{m_e}c^2), \qquad \beta = \frac{\sqrt{E_{\mathrm{e}}(E_{\mathrm{e}} + 2\mathrm{m_e}c^2)}}{E_{\mathrm{e}} + \mathrm{m_e}c^2}. \quad (2.25)$$

Strictly speaking, the DCS (2.24) is adequate only for ionization of the K-shell by high-energy photons. Nevertheless, in many practical simulations no appreciable errors are introduced when Sauter's distribution is used to describe any photoionization event, irrespective of the atomic shell and the photon energy. The main reason is that the emitted photoelectron immediately starts to interact with the medium, and its direction of movement is strongly altered after travelling a path length much shorter than the photon mean free path. On the other hand, when the photon energy exceeds the K-edge, most of the ionizations occur in the K-shell and then the Sauter distribution represents a good approximation.

Introducing the variable $\nu = 1 - \cos\theta_{\mathrm{e}}$, the angular distribution of photoelectrons can be expressed in the form

$$p(\nu) = (2-\nu)\left[\frac{1}{A+\nu} + \frac{1}{2}\beta\gamma(\gamma-1)(\gamma-2)\right]\frac{\nu}{(A+\nu)^3}, \qquad A = \frac{1}{\beta} - 1, \quad (2.26)$$

apart from a normalization constant. Random sampling of ν from this distribution can be performed analytically. To this end, $p(\nu)$ can be factorized in the form

$$p(\nu) = g(\nu)\pi(\nu) \quad (2.27)$$

with

$$g(\nu) = (2-\nu)\left[\frac{1}{A+\nu} + \frac{1}{2}\beta\gamma(\gamma-1)(\gamma-2)\right] \quad (2.28)$$

and

$$\pi(\nu) = \frac{A(A+2)^2}{2}\frac{\nu}{(A+\nu)^3}. \quad (2.29)$$

The variable ν takes values in the interval (0,2), where the function $g(\nu)$ is definite positive and attains its maximum value at $\nu = 0$, while the function $\pi(\nu)$ is positive and normalized to unity. Random values from the probability distribution $\pi(\nu)$ are generated by means of the sampling formula (inverse transform method, see section 1.2.2)

$$\int_0^{\nu} \pi(\nu')\,\mathrm{d}\nu' = \xi, \quad (2.30)$$

which can be solved analytically to give

$$\nu = \frac{2A}{(A+2)^2 - 4\xi}\left[2\xi + (A+2)\xi^{1/2}\right]. \quad (2.31)$$

Therefore, random sampling from Sauter's distribution can be performed by the rejection method (see section 1.2.4) as follows:

(i) Generate ν from $\pi(\nu)$ by using eq. (2.31).

(ii) Generate a random number ξ.

(iii) If $\xi g(0) > g(\nu)$, go to step (i).

(iv) Deliver $\cos\theta_e = 1 - \nu$.

The efficiency of this algorithm is ~ 0.33 at low energies and increases slowly with E_e; for $E_e = 1$ MeV, the efficiency is 0.4. As photoelectric absorption occurs at most once in each photon history, this small sampling efficiency does not slow down the simulation significantly.

2.3 Incoherent (Compton) scattering

In Compton scattering, a photon of energy E interacts with an atomic electron, which absorbs it and re-emits a secondary (Compton) photon of energy E' in the direction $\Omega = (\theta, \phi)$ relative to the direction of the original photon. In PENELOPE, Compton scattering events are described by means of the cross section obtained from the relativistic impulse approximation (Ribberfors, 1983). Contributions from different atomic electron shells are considered separately. After a Compton interaction with the i-th shell, the active target electron is ejected to a free state with kinetic energy $E_e = E - E' - U_i > 0$, where U_i is the ionization energy of the considered shell, and the residual atom is left in an excited state with a vacancy in the i-th shell.

In the case of scattering by free electrons at rest, the conservation of energy and momentum implies the following relation between the energy E' of the scattered (Compton) photon and the scattering angle θ [cf. eq. (A.19)]

$$E' \equiv \frac{E}{1+\kappa(1-\cos\theta)} \equiv E_C, \tag{2.32}$$

where $\kappa = E/m_e c^2$, as before. The DCS for Compton scattering by a free electron at rest is given by the familiar Klein-Nishina formula,

$$\frac{d\sigma_{Co}^{KN}}{d\Omega} = \frac{r_e^2}{2}\left(\frac{E_C}{E}\right)^2\left(\frac{E_C}{E} + \frac{E}{E_C} - \sin^2\theta\right). \tag{2.33}$$

Although this simple DCS was generally used in old Monte Carlo transport codes, it represents only a rough approximation for the Compton interactions of photons with atoms. In reality, atomic electrons are not at rest, but move with a certain momentum distribution, which gives rise to the so-called Doppler broadening of the Compton line. Moreover, transitions of bound electrons are allowed only if the energy transfer $E - E'$ is larger than the ionization energy U_i of the active shell (binding effect).

The impulse approximation accounts for Doppler broadening and binding effects in a natural, and relatively simple, way. The DCS is obtained by considering that

electrons in the i-th shell move with a momentum distribution $\rho_i(\mathbf{p})$. For an electron in an orbital $\psi_i(\mathbf{r})$, $\rho_i(\mathbf{p}) \equiv |\psi_i(\mathbf{p})|^2$, where $\psi_i(\mathbf{p})$ is the wave function in the momentum representation. The DCS for Compton scattering by an electron with momentum $\mathbf{p}$ is derived from the Klein-Nishina formula by applying a Lorentz transformation with velocity $\mathbf{v}$ equal to that of the moving target electron. The impulse approximation to the Compton DCS (per electron) of the considered shell is obtained by averaging over the momentum distribution $\rho_i(\mathbf{p})$.

After some manipulations, the Compton DCS of an electron in the i-th shell can be expressed as [eq. (21) in Brusa et al., 1996]

$$\frac{\mathrm{d}^2\sigma_{\mathrm{Co},i}}{\mathrm{d}E'\mathrm{d}\Omega} = \frac{r_\mathrm{e}^2}{2}\left(\frac{E_\mathrm{C}}{E}\right)^2\left(\frac{E_\mathrm{C}}{E}+\frac{E}{E_\mathrm{C}}-\sin^2\theta\right)F(p_z)\,J_i(p_z)\,\frac{\mathrm{d}p_z}{\mathrm{d}E'}, \tag{2.34}$$

where r_e is the classical electron radius. E_C is the energy of the Compton line, defined by eq. (2.32), i.e. the energy of photons scattered in the direction θ by *free electrons at rest*. The momentum transfer vector is given by $\mathbf{q} \equiv \hbar\mathbf{k} - \hbar\mathbf{k}'$, where $\hbar\mathbf{k}$ and $\hbar\mathbf{k}'$ are the momenta of the incident and scattered photons; its magnitude is

$$q = \frac{1}{c}\sqrt{E^2 + E'^2 - 2EE'\cos\theta}. \tag{2.35}$$

The quantity p_z is the projection of the initial momentum $\mathbf{p}$ of the electron on the direction of the scattering vector $\hbar\mathbf{k}' - \hbar\mathbf{k} = -\mathbf{q}$; it is given by[2]

$$p_z \equiv -\frac{\mathbf{p}\cdot\mathbf{q}}{q} = \frac{EE'(1-\cos\theta) - \mathrm{m_e}c^2(E-E')}{c^2 q} \tag{2.36}$$

or, equivalently,

$$\frac{p_z}{\mathrm{m_e}c} = \frac{E(E'-E_\mathrm{C})}{E_\mathrm{C}\,cq}. \tag{2.37}$$

Notice that $p_z = 0$ for $E' = E_\mathrm{C}$. Moreover,

$$\frac{\mathrm{d}p_z}{\mathrm{d}E'} = \frac{\mathrm{m_e}c}{cq}\left(\frac{E}{E_\mathrm{C}} + \frac{E\cos\theta - E'}{cq}\,\frac{p_z}{\mathrm{m_e}c}\right). \tag{2.38}$$

The function $J_i(p_z)$ in eq. (2.34) is the one-electron Compton profile of the active shell, which is defined as

$$J_i(p_z) \equiv \iint \rho_i(\mathbf{p})\,\mathrm{d}p_x\,\mathrm{d}p_y, \tag{2.39}$$

where $\rho_i(\mathbf{p})$ is the electron momentum distribution. That is, $J_i(p_z)\,\mathrm{d}p_z$ gives the probability that the component of the electron momentum in the z-direction is in the interval $(p_z, p_z + \mathrm{d}p_z)$. Notice that the normalization

$$\int_{-\infty}^{\infty} J_i(p_z)\,\mathrm{d}p_z = 1 \tag{2.40}$$

[2]The expression (2.36) contains an approximation, the exact relation is obtained by replacing the electron rest energy $\mathrm{m_e}c^2$ in the numerator by the electron initial total energy, $\sqrt{(\mathrm{m_e}c^2)^2 + (cp)^2}$.

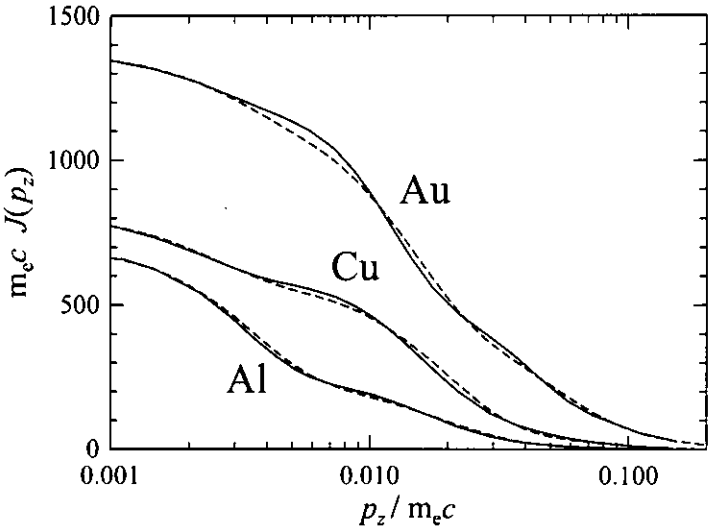


Figure 2.6: Atomic Compton profiles ($p_z > 0$) for aluminium, copper and gold. The continuous curves are numerical Hartree-Fock profiles tabulated by Biggs et al. (1975). The dashed curves represent the analytical profiles defined by eq. (2.57). (Adapted from Brusa et al., 1996.)

is assumed. In the Hartree-Fock approximation for closed-shell configurations, the momentum distribution of the electrons in an atomic shell, obtained by adding the contributions of the orbitals in that shell, is isotropic. For an isotropic distribution, expression (2.39) simplifies to

$$J_i(p_z) = 2\pi \int_{|p_z|}^{\infty} p\, \rho_i(p)\, \mathrm{d}p. \tag{2.41}$$

The atomic Compton profile is given by

$$J(p_z) = \sum_i f_i\, J_i(p_z), \tag{2.42}$$

where f_i is the number of electrons in the i-th shell and $J_i(p_z)$ is the one-electron profile of this shell. The functions $J(p_z)$ and $J_i(p_z)$ are both bell-shaped and symmetrical about $p_z = 0$ (see fig. 2.6). Extensive tables of Hartree-Fock Compton profiles for the elements have been published by Biggs et al. (1975). These numerical profiles are adequate for bound electron shells. In the case of conductors, the one-electron Compton profile for conduction electrons may be estimated by assuming that these form a free-electron gas with ρ_e electrons per unit volume. The one-electron profile for this system is (see e.g. Cooper, 1971)

$$J_i^{\mathrm{feg}}(p_z) = \frac{3}{4p_{\mathrm{F}}}\left(1 - \frac{p_z^2}{p_{\mathrm{F}}^2}\right)\Theta(p_{\mathrm{F}} - |p_z|), \qquad J_i^{\mathrm{feg}}(0) = \frac{3}{4p_{\mathrm{F}}}, \tag{2.43}$$

where $p_{\mathrm{F}} \equiv \hbar(3\pi^2\rho_e)^{1/3}$ is the Fermi momentum. For scattering in a compound material,

 3482

the molecular Compton profile is obtained as the sum of atomic profiles of the atoms in a molecule (additivity rule).

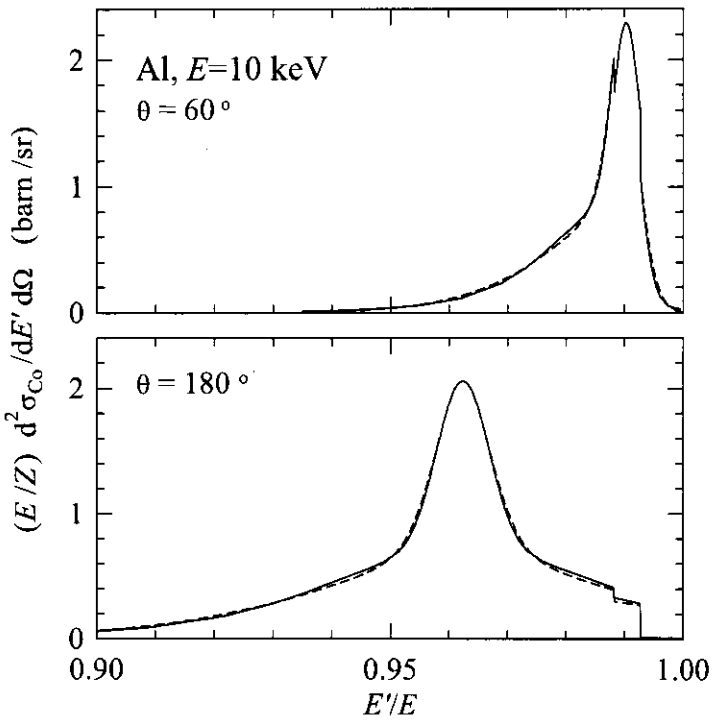


Figure 2.7: DCS for Compton scattering of 10 keV photons by aluminium atoms at the indicated scattering angles. The continuous curves represent the DCS (2.46) calculated using the Hartree-Fock Compton profile (Biggs et al., 1975). The dashed curves are results from eq. (2.46) with the analytical profiles given by eq. (2.57). (Adapted from Brusa et al., 1996.)

The factor $F(p_z)$ in eq. (2.34) is approximately given by

$$F(p_z) \simeq 1 + \frac{cq_C}{E}\left(1 + \frac{E_C(E_C - E\cos\theta)}{(cq_C)^2}\right)\frac{p_z}{m_e c}, \tag{2.44}$$

where q_C is the momentum transfer associated with the energy $E' = E_C$ of the Compton line,

$$q_C \equiv \frac{1}{c}\sqrt{E^2 + E_C^2 - 2EE_C\cos\theta}. \tag{2.45}$$

Expression (2.44) is accurate only for small $|p_z|$-values. For large $|p_z|$, $J_i(p_z)$ tends to zero and the factor $F(p_z)$ has no effect on the DCS. We use the values given by expression (2.44) only for $|p_z| < 0.2 m_e c$ and take $F(\pm|p_z|) = F(\pm 0.2 m_e c)$ for $|p_z| > 0.2 m_e c$. Owing to the approximations introduced, negative values of F may be obtained for large $|p_z|$; in this case, we must set $F = 0$.

We can now introduce the effect of electron binding: Compton excitations are allowed only if the target electron is promoted to a free state, i.e. if the energy transfer $E - E'$

is larger than the ionization energy U_i of the active shell. Therefore the atomic DCS, including Doppler broadening and binding effects, is given by

$$\frac{\mathrm{d}^2\sigma_{\mathrm{Co}}}{\mathrm{d}E'\mathrm{d}\Omega} = \frac{r_\mathrm{e}^2}{2}\left(\frac{E_\mathrm{C}}{E}\right)^2\left(\frac{E_\mathrm{C}}{E}+\frac{E}{E_\mathrm{C}}-\sin^2\theta\right) \times F(p_z)\left(\sum_i f_i\, J_i(p_z)\,\Theta(E-E'-U_i)\right)\frac{\mathrm{d}p_z}{\mathrm{d}E'}, \tag{2.46}$$

where $\Theta(x)$ (= 1 if $x > 0$, = 0 otherwise) is the Heaviside step function. In the calculations we use the ionization energies U_i given by Lederer and Shirley (1978), fig. 2.4. The DCS for scattering of 10 keV photons by aluminium atoms is displayed in fig. 2.7, for θ = 60 and 180 degrees, as a function of the fractional energy of the emerging photon. The DCS for a given scattering angle has a maximum at $E' = E_\mathrm{C}$; its shape resembles that of the atomic Compton profile, except for the occurrence of edges at $E' = E - U_i$.

In the case of scattering by free electrons at rest we have $U_i = 0$ (no binding) and $J_i(p_z) = \delta(p_z)$ (no Doppler broadening). Moreover, from eq. (2.37) $E' = E_\mathrm{C}$, so that photons scattered through an angle θ have energy E_C. Integration of the DCS, eq. (2.46), over E' then yields the familiar Klein-Nishina cross section,

$$\frac{\mathrm{d}\sigma_{\mathrm{Co}}^{\mathrm{KN}}}{\mathrm{d}\Omega} = Z\,\frac{r_\mathrm{e}^2}{2}\left(\frac{E_\mathrm{C}}{E}\right)^2\left(\frac{E_\mathrm{C}}{E}+\frac{E}{E_\mathrm{C}}-\sin^2\theta\right), \tag{2.47}$$

for the Z atomic electrons [cf. eq. (2.33)]. For energies of the order of a few MeV and larger, Doppler broadening and binding effects are relatively small and the free-electron theory yields results practically equivalent to those of the impulse approximation.

The angular distribution of scattered photons is given by the directional DCS,

$$\frac{\mathrm{d}\sigma_{\mathrm{Co}}}{\mathrm{d}\Omega} = \int\frac{\mathrm{d}^2\sigma_{\mathrm{Co}}}{\mathrm{d}E'\mathrm{d}\Omega}\,\mathrm{d}E' = \frac{r_\mathrm{e}^2}{2}\left(\frac{E_\mathrm{C}}{E}\right)^2\left(\frac{E_\mathrm{C}}{E}+\frac{E}{E_\mathrm{C}}-\sin^2\theta\right) \times\sum_i f_i\,\Theta(E-U_i)\int_{-\infty}^{p_{i,\max}} F(p_z)J_i(p_z)\,\mathrm{d}p_z, \tag{2.48}$$

where $p_{i,\max}$ is the highest p_z-value for which an electron in the i-th shell can be excited. It is obtained from eq. (2.36) by setting $E' = E - U_i$,

$$p_{i,\max}(E,\theta) = \frac{E(E-U_i)(1-\cos\theta) - \mathrm{m_e}c^2U_i}{c\sqrt{2E(E-U_i)(1-\cos\theta)+U_i^2}}. \tag{2.49}$$

Except for energies just above the shell ionization threshold, the function $F(p_z)$ in the integral can be replaced by unity, since $p_z J_i(p_z)$ is an odd function and its integral is close to zero, i.e.

$$\int_{-\infty}^{p_{i,\max}} F(p_z)J_i(p_z)\,\mathrm{d}p_z \simeq n_i(p_{i,\max}), \tag{2.50}$$

where

$$n_i(p_z) \equiv \int_{-\infty}^{p_z} J_i(p'_z)\,\mathrm{d}p'_z. \tag{2.51}$$

Notice that $n_i(p_z)$ is a monotonously increasing function of p_z, which varies from 0 at $p_z = -\infty$ to unity at $p_z = \infty$; the quantity $n_i(p_{i,\max})$ represents the fraction of electrons in the i-th shell that can be effectively excited in a Compton interaction. We can then write

$$\frac{\mathrm{d}\sigma_{\mathrm{Co}}}{\mathrm{d}\Omega} \simeq \frac{r_e^2}{2}\left(\frac{E_C}{E}\right)^2\left(\frac{E_C}{E} + \frac{E}{E_C} - \sin^2\theta\right) S(E,\theta). \tag{2.52}$$

The function

$$S(E,\theta) = \sum_i f_i\,\Theta(E - U_i)\,n_i(p_{i,\max}) \tag{2.53}$$

can be identified with the incoherent scattering function in the impulse approximation (see e.g. Ribberfors and Berggren, 1982). The total cross section can then be obtained as

$$\sigma_{\mathrm{Co}} = 2\pi \int \frac{\mathrm{d}\sigma_{\mathrm{Co}}}{\mathrm{d}\Omega}\,\mathrm{d}(\cos\theta). \tag{2.54}$$

For comparison purposes, and also to calculate the energy deposition, it is useful to consider the cross section differential in only the energy of the scattered photon,

$$\frac{\mathrm{d}\sigma_{\mathrm{Co}}}{\mathrm{d}E'} \equiv \int \frac{\mathrm{d}^2\sigma_{\mathrm{Co}}}{\mathrm{d}E'\mathrm{d}\Omega}\,\mathrm{d}\Omega. \tag{2.55}$$

In the case of scattering by free electrons at rest, $E' = E_C$ and the Klein-Nishina formula (2.47) gives the following expression for the energy DCS,

$$\begin{aligned}\frac{\mathrm{d}\sigma_{\mathrm{Co}}^{\mathrm{KN}}}{\mathrm{d}E'} &= 2\pi\,\frac{\mathrm{d}\sigma_{\mathrm{Co}}^{\mathrm{KN}}}{\mathrm{d}\Omega}\,\frac{\mathrm{d}(\cos\theta)}{\mathrm{d}E_C} \\ &= \frac{\pi r_e^2}{E}\,\kappa^{-3}\left(\frac{E^2}{E'^2} + \frac{(\kappa^2 - 2\kappa - 2)E}{E'} + (2\kappa + 1) + \frac{\kappa^2 E'}{E}\right).\end{aligned} \tag{2.56}$$

Fig. 2.8 displays the energy DCSs obtained from this formula and from the impulse approximation for scattering of high-energy ($E > U_i$) photons by aluminium and gold atoms. These results show clearly the differences between the physics of the impulse approximation and the cruder free-electron approximation. The most conspicuous feature of the impulse approximation DCS is the absence of a threshold energy, which is a direct manifestation of the Doppler broadening. For relatively small energy transfers ($E' \sim E$) the Klein-Nishina DCS increases with the energy of the scattered photon, whereas the energy DCS obtained from the impulse approximation vanishes at $E' = E$ due to the effect of binding, which also causes the characteristic edge structure, similar to that of the photoelectric cross section (see fig. 2.8).

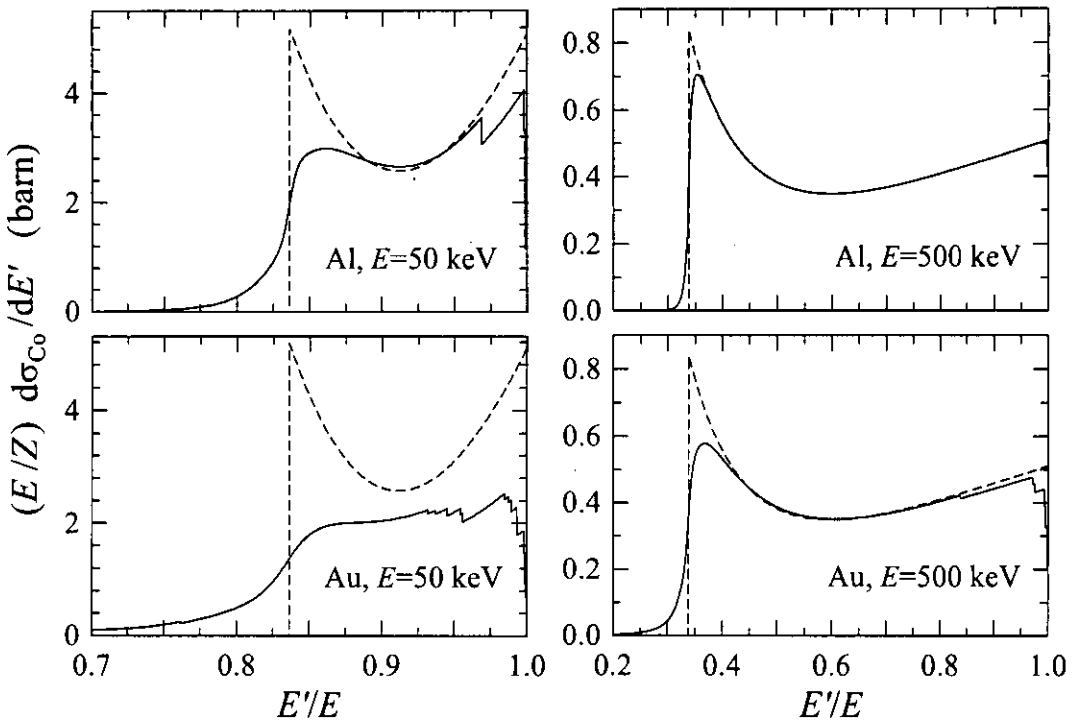


Figure 2.8: Energy DCSs for Compton scattering of 50 and 500 keV photons by aluminium and gold atoms. The continuous curves represent the DCS (2.55), computed using the analytical Compton profiles (2.57). The dashed curves are obtained from the Klein-Nishina formula (2.56), i.e. assuming that the atomic electrons are free and at rest.

2.3.1 Analytical Compton profiles

In order to minimize the required numerical information and to simplify the random sampling, we use approximate one-electron profiles of the form

$$J_i^{\mathrm{A}}(p_z) = J_{i,0}\,\frac{n d_2}{2}\left(d_1 + d_2 J_{i,0}\,|p_z|\right)^{n-1} \exp\left[d_1^n - \left(d_1 + d_2 J_{i,0}\,|p_z|\right)^n\right] \qquad (2.57)$$

with

$$n = 2, \qquad d_1 = \left(\frac{n-1}{n}\right)^{1/n} = \sqrt{\frac{1}{2}}, \qquad d_2 = \frac{2}{n}\,d_1^{1-n} = \sqrt{2}.$$

The quantity $J_{i,0} \equiv J_i(0)$ is the value of the profile at $p_z = 0$ obtained from the Hartree-Fock orbital (Biggs et al., 1975). $J_i(0)$ is tabulated in the file `PDATCONF.TAB` for all shells of the elements $Z = 1$ to 92. Notice that $J_i^{\mathrm{A}}(p_z)$ is normalized according to eq.

(2.40). With the profiles (2.57),

$$n_i^{\mathrm{A}}(p_z) \equiv \int_{-\infty}^{p_z} J_i^{\mathrm{A}}(p_z')\,\mathrm{d}p_z' = \begin{cases} \frac{1}{2}\exp\left[d_1^2 - \left(d_1 - d_2 J_{i,0}\,p_z\right)^2\right] & \text{if } p_z < 0, \\ 1 - \frac{1}{2}\exp\left[d_1^2 - \left(d_1 + d_2 J_{i,0}\,p_z\right)^2\right] & \text{if } p_z > 0. \end{cases} \tag{2.58}$$

Thus, the incoherent scattering function (2.53) can be expressed analytically and the integral (2.54) evaluated very quickly with the aid of function SUMGA (appendix B). On the other hand, the sampling equation $n_i^{\mathrm{A}}(p_z) \equiv \xi n_i^{\mathrm{A}}(p_{i,\max})$ (see section 1.2.2) can be solved analytically,

$$p_z = \begin{cases} \dfrac{1}{d_2 J_{i,0}}\left[d_1 - \left({d_1}^2 - \ln 2A\right)^{1/2}\right] & \text{if } A < \frac{1}{2}, \\ \dfrac{1}{d_2 J_{i,0}}\left[\left({d_1}^2 - \ln 2(1-A)\right)^{1/2} - d_1\right] & \text{if } A > \frac{1}{2}, \end{cases} \tag{2.59}$$

where $A \equiv \xi n_i^{\mathrm{A}}(p_{i,\max})$. Atomic Compton profiles obtained from the approximation given by eq. (2.57) are accurate for small p_z and oscillate about the Hartree-Fock values for intermediate momenta (see fig. 2.6). The relative differences are normally less than 5%, except for large momenta for which $J(p_z)$ is very small. Similar differences are found between the DCS computed from Hartree-Fock and analytical Compton profiles (see fig. 2.7). For most applications (e.g. studies of detector response, dosimetry, radiotherapy, etc.), the effect of these differences on the simulation results is not important. The impulse approximation with the analytical one-electron profiles (2.57) then provides a conveniently simple method to introduce Doppler broadening and binding effects in the simulation of Compton scattering.

2.3.2 Simulation of incoherent scattering events

Compton events are simulated on the basis of the DCS given by eq. (2.46) with the analytical Compton profiles (2.57). The sampling algorithm adopted here is due to Brusa et al. (1996). It is similar to the one described by Namito et al. (1994), but has a higher efficiency.

The PDF of the polar deflection $\cos\theta$ and the energy E' of the scattered photon is given by (apart from normalization constants, which are irrelevant here)

$$P_{\mathrm{Co}}(\cos\theta, E') = \left(\frac{E_{\mathrm{C}}}{E}\right)^2 \left(\frac{E_{\mathrm{C}}}{E} + \frac{E}{E_{\mathrm{C}}} - \sin^2\theta\right) \times F(p_z) \left(\sum_i f_i\, J_i(p_z)\,\Theta(E - E' - U_i)\right) \frac{\mathrm{d}p_z}{\mathrm{d}E'}. \tag{2.60}$$

Integration of expression (2.60) over E', using the approximation (2.50), yields the PDF of the polar deflection

$$P_\theta(\cos\theta) = \left(\frac{E_{\mathrm{C}}}{E}\right)^2 \left(\frac{E_{\mathrm{C}}}{E} + \frac{E}{E_{\mathrm{C}}} - \sin^2\theta\right) S(E,\theta), \tag{2.61}$$

where $S(E,\theta)$ is the incoherent scattering function, eq. (2.53).

Random values of $\cos\theta$ from the PDF (2.61) can be generated by using the following algorithm (Baró et al., 1994a). Let us introduce the quantity

$$\tau \equiv \frac{E_C}{E} = \frac{1}{1+\kappa(1-\cos\theta)}. \tag{2.62}$$

The minimum and maximum values of τ are

$$\tau_{\min} = \frac{1}{1+2\kappa} \qquad \text{and} \qquad \tau_{\max} = 1, \tag{2.63}$$

which correspond to backward ($\theta = \pi$) and forward ($\theta = 0$) scattering, respectively. The PDF of this variable is (again ignoring normalization constants)

$$P_\tau(\tau) = P_\theta(\cos\theta)\frac{\mathrm{d}(\cos\theta)}{\mathrm{d}\tau} = \left(\frac{1}{\tau^2} + \frac{\kappa^2-2\kappa-2}{\tau} + (2\kappa+1) + \kappa^2\tau\right) S(E,\theta). \tag{2.64}$$

This distribution can be rewritten in the form (Nelson et al., 1985)

$$P_\tau(\tau) = [a_1 P_1(\tau) + a_2 P_2(\tau)]\, T(\cos\theta), \tag{2.65}$$

where

$$a_1 = \ln(1+2\kappa), \qquad a_2 = \frac{2\kappa(1+\kappa)}{(1+2\kappa)^2}, \tag{2.66}$$

$$P_1(\tau) = \frac{1}{\ln(1+2\kappa)}\frac{1}{\tau}, \qquad P_2(\tau) = \frac{(1+2\kappa)^2}{2\kappa(1+\kappa)}\tau \tag{2.67}$$

and

$$T(\cos\theta) = \left\{1 - \frac{(1-\tau)\,[(2\kappa+1)\tau - 1]}{\kappa^2\tau(1+\tau^2)}\right\} \frac{S(E,\theta)}{S(E,\theta=\pi)}. \tag{2.68}$$

The function in braces is positive, it equals 1 at the end points of the interval $(\tau_{\min},1)$, and is less than unity inside this interval. Moreover, the ratio of incoherent scattering functions is also less than unity for any value of $\theta < \pi$. Hence, the function $T(\cos\theta)$ is a valid rejection function. The functions $P_i(\tau)$ $(i = 1,2)$ are normalized PDFs in the interval $(\tau_{\min},1)$, which can be easily sampled by using the inverse transform method. The generation of random values of τ according to the PDF given by eq. (2.64) can then be performed by combining the composition and rejection methods (section 1.2). The algorithm to sample $\cos\theta$ proceeds as follows:

(i) Sample a value of the integer i (=1, 2) according to the point probabilities

$$\pi(1) = \frac{a_1}{a_1+a_2} \qquad \text{and} \qquad \pi(2) = \frac{a_2}{a_1+a_2}. \tag{2.69}$$

(ii) Sample τ from $P_i(\tau)$ using the sampling formulae

$$\tau = \begin{cases} \tau_{\min}^{\xi} & \text{if } i = 1, \\ [\tau_{\min}^2 + \xi\,(1 - \tau_{\min}^2)]^{1/2} & \text{if } i = 2, \end{cases} \tag{2.70}$$

which can be easily derived by the inverse transform method (section 1.2.2).

(iii) Determine $\cos\theta$ using eq. (2.62),

$$\cos\theta = 1 - \frac{1-\tau}{\kappa\tau}, \tag{2.71}$$

and compute the quantities $p_{i,\max}(E,\theta)$, eq. (2.49), and

$$S(E,\theta) = \sum_i f_i\,\Theta(E - U_i)\,n_i^{\mathrm{A}}(p_{i,\max}). \tag{2.72}$$

(iv) Generate a new random number ξ.

(v) If $\xi > T(\cos\theta)$, go to step (i).

(vi) Deliver $\cos\theta$.

The efficiency of this algorithm, i.e. the probability of accepting a generated $\cos\theta$-value, increases monotonically with photon energy and is nearly independent of Z; typical values are 35%, 80% and 95% for $E = 1$ keV, 1 MeV and 10 MeV, respectively.

Once the direction of the emerging photon has been set, the active electron shell i is selected with relative probability equal to $Z_i\,\Theta(E - U_i)\,n_i^{\mathrm{A}}(p_{i,\max}(E,\theta))$. A random value of p_z is generated from the analytical Compton profile (2.57) using the sampling formula (2.59). If p_z is less than $-\mathrm{m_e}c$, it is rejected and a new shell and a p_z-value are sampled[3]. Finally, the factor $F(p_z)$ in the PDF (2.46) is accounted for by means of a rejection procedure. It should be noted that the approximation $F \simeq 1$ is valid only when the DCS is integrated over E'; otherwise the complete expression (2.44) must be used. Let $F_{\max}$ denote the maximum value of $F(p_z)$, which occurs at $p_z = 0.2\mathrm{m_e}c$ or $-0.2\mathrm{m_e}c$; a random number ξ is generated and the value p_z is accepted if $\xi F_{\max} < F(p_z)$, otherwise the process of selecting a shell and a p_z-value is reinitiated. The energy E' of the emerging photon is then calculated from eq. (2.36), which gives

$$E' = E\,\frac{\tau}{1 - t\tau^2}\left[(1 - t\tau\cos\theta) + \mathrm{sign}(p_z)\sqrt{(1 - t\tau\cos\theta)^2 - (1 - t\tau^2)(1-t)}\right], \tag{2.73}$$

where

$$t \equiv (p_z/\mathrm{m_e}c)^2 \qquad \text{and} \qquad \mathrm{sign}(p_z) \equiv p_z/|p_z|. \tag{2.74}$$

For photons with energy larger than 5 MeV, for which Doppler broadening is negligible, we set $E' = E_{\mathrm{C}}$ (which amounts to assuming that $p_z = 0$). In this case, the active

[3]Notice that, due to the approximation introduced in eq. (2.36), a value $p_z < -\mathrm{m_e}c$ would yield a negative energy for the scattered photon.

electron shell i is sampled with relative probability Z_i and binding effects are accounted for by simply rejecting E'-values such that $E - E' < U_i$.

The azimuthal scattering angle ϕ of the photon is sampled uniformly in the interval $(0, 2\pi)$. We assume that the Compton electron is emitted with energy $E_e = E - E' - U_i$ in the direction of the momentum transfer vector $\mathbf{q} = \hbar\mathbf{k} - \hbar\mathbf{k}'$, with polar angle θ_e and azimuthal angle $\phi_e = \phi + \pi$, relative to the direction of the incident photon. $\cos\theta_e$ is given by

$$\cos\theta_e = \frac{E - E'\cos\theta}{\sqrt{E^2 + E'^2 - 2EE'\cos\theta}}. \tag{2.75}$$

When $E' = E_C$, this expression simplifies to

$$\cos\theta_e = \frac{E + m_e c^2}{E}\left(\frac{E - E_C}{2m_e c^2 + E - E_C}\right)^{1/2}, \tag{2.76}$$

which coincides with the result (A.20). Since the active electron shell is known, characteristic x-rays and electrons emitted in the de-excitation of the ionized atom can also be followed. This is important, for instance, to account for escape peaks in scintillation or solid state detectors

Table 2.1: Average number n_r of random numbers ξ needed to simulate a single incoherent scattering event for photons with energy E in aluminium, silver and gold.

E (eV)	Al	Ag	Au
10^3	16.6	11.9	13.4
10^4	11.0	11.4	11.5
10^5	9.5	9.8	10.0
10^6	8.2	8.2	8.3
10^7	7.5	7.5	7.5

As a measure of the efficiency of the sampling algorithm, we may consider the average number n_r of random numbers ξ required to simulate an incoherent scattering event. n_r is practically independent of the atomic number and decreases with photon energy (see table 2.1). The increase of n_r at low energies stems from the loss of efficiency of the algorithm used to sample $\cos\theta$. Although the simulation of incoherent events becomes more laborious as the photon energy decreases, this has only a small influence on the speed of practical photon transport simulations since low-energy photons interact predominantly via photoelectric absorption (see fig. 2.10 below).

2.4 Electron-positron pair production

Electron-positron pairs can be created by absorption of a photon in the vicinity of a massive particle, a nucleus or an electron, which absorbs energy and momentum so that

 3490

these two quantities are conserved. The threshold energy for pair production in the field of a nucleus (assumed of infinite mass) is $2m_ec^2$. When pair production occurs in the field of an electron, the target electron recoils after the event with appreciable kinetic energy; the process is known as "triplet production" because it causes three visible tracks when observed, e.g. in a cloud chamber. If the target electron is at rest, triplet production is only possible for photons with energy larger than $4m_ec^2$.

For the simulation of pair production events in the field of an atom of atomic number Z, we shall use the following semiempirical model (Baró et al., 1994a). Our starting point is the high-energy DCS for arbitrary screening, which was derived by Bethe and Heitler from the Born approximation (Motz et al., 1969; Tsai, 1974). The Bethe-Heitler DCS for a photon of energy E to create an electron-positron pair, in which the electron has a kinetic energy $E_- = \epsilon E - m_ec^2$, can be expressed as (Tsai, 1974)

$$\frac{d\sigma_{pp}^{(BH)}}{d\epsilon} = r_e^2\alpha Z[Z+\eta]\left\{\left[\epsilon^2+(1-\epsilon)^2\right](\Phi_1-4f_C)+\frac{2}{3}\epsilon(1-\epsilon)(\Phi_2-4f_C)\right\}. \quad (2.77)$$

Notice that the "reduced energy" $\epsilon = (E_- + m_ec^2)/E$ is the fraction of the photon energy that is taken away by the electron. The screening functions Φ_1 and Φ_2 are given by integrals that involve the atomic form factor and, therefore, must be computed numerically when a realistic form factor is adopted (e.g. the analytical one described in section 2.1). To obtain approximate analytical expressions for these functions, we shall assume that the Coulomb field of the nucleus is exponentially screened by the atomic electrons (Schiff, 1968; Tsai, 1974), i.e. the electrostatic potential of the atom is assumed to be (Wentzel model)

$$\varphi_W(r) = \frac{Ze}{r}\exp(-r/R), \quad (2.78)$$

with the screening radius R considered as an adjustable parameter (see below). The corresponding atomic electron density is obtained from Poisson's equation,

$$\rho_W(r) = \frac{1}{4\pi e}\nabla^2\varphi(r) = \frac{1}{4\pi e}\frac{1}{r}\frac{d^2}{dr^2}[r\varphi(r)] = \frac{Z}{4\pi R^2 r}\exp(-r/R), \quad (2.79)$$

and the atomic form factor is

$$F_W(q,Z) = 4\pi\int_0^\infty \rho_W(r)\frac{\sin(qr/\hbar)}{qr/\hbar}r^2\,dr = \frac{Z}{1+(Rq/\hbar)^2}. \quad (2.80)$$

The screening functions for this particular form factor take the following analytical expressions (Tsai, 1974)

$$\begin{aligned}\Phi_1 &= 2-2\ln(1+b^2)-4b\arctan(b^{-1})+4\ln(Rm_ec/\hbar)\\ \Phi_2 &= \frac{4}{3}-2\ln(1+b^2)+2b^2\left[4-4b\arctan(b^{-1})-3\ln(1+b^{-2})\right]\\ &\quad +4\ln(Rm_ec/\hbar),\end{aligned} \quad (2.81)$$

where

$$b = \frac{Rm_ec}{\hbar}\frac{1}{2\kappa}\frac{1}{\epsilon(1-\epsilon)}. \quad (2.82)$$

The quantity η in eq. (2.77) accounts for pair production in the field of the atomic electrons (triplet production), which is considered in detail by Hubbell et al. (1980) and Tsai (1974). In order to simplify the calculations, the dependence of the triplet cross section on the electron reduced energy, ϵ, is assumed to be the same as that of the pair cross section. The function f_C in (2.77) is the high-energy Coulomb correction of Davies, Bethe and Maximon (1954) given by

$$f_C(Z) = a^2\left[(1+a^2)^{-1} + 0.202059 - 0.03693a^2 + 0.00835a^4 - 0.00201a^6 + 0.00049a^8 - 0.00012a^{10} + 0.00003a^{12}\right], \quad (2.83)$$

with $a = \alpha Z$. The total atomic cross section for pair (and triplet) production is obtained as

$$\sigma_{pp}^{(BH)} = \int_{\epsilon_{min}}^{\epsilon_{max}} \frac{d\sigma_{pp}^{(BH)}}{d\epsilon}\, d\epsilon, \quad (2.84)$$

where

$$\epsilon_{min} = m_e c^2/E = \kappa^{-1} \qquad \text{and} \qquad \epsilon_{max} = 1 - m_e c^2/E = 1 - \kappa^{-1}. \quad (2.85)$$

Extensive tables of pair production total cross sections, evaluated by combining different theoretical approximations, have been published by Hubbell et al. (1980). These tables give the separate contributions of pair production in the field of the nucleus and in that of the atomic electrons for $Z = 1$ to 100 and for photon energies from threshold up to 10^5 MeV. Following Salvat and Fernández-Varea (1992), the screening radius R has been determined by requiring that eq. (2.77) with $\eta = 0$ exactly reproduces the total cross sections given by Hubbell et al. (1980) for pair production in the nuclear field by 10^5 MeV photons (after exclusion of radiative corrections, which only amount to $\sim 1\%$ of the total cross section). The screening radii for $Z = 1$–92 obtained in this way are given in table 2.2.

Actually, the triplet contribution, η, varies with the photon energy. It increases monotonically from zero at $E \simeq 4m_e c^2$ and reaches a saturation value, η_∞, at high energies. It can be obtained, for all elements and energies up to 10^5 MeV, as

$$\eta(E) = Z\sigma_{triplet}^{HGO}(E)/\sigma_{pair}^{HGO}(E), \quad (2.86)$$

where σ_{pair}^{HGO} and $\sigma_{triplet}^{HGO}$ are the total pair and triplet production cross sections given by Hubbell et al. (1980). At 10^5 MeV, the high-energy limit is reached, i.e.

$$\eta_\infty \simeq Z\sigma_{triplet}^{HGO}(10^5\,\text{MeV})/\sigma_{pair}^{HGO}(10^5\,\text{MeV}). \quad (2.87)$$

The values of η_∞ for the elements $Z = 1$–92 are given in table 2.2. The average dependence of η on the photon energy is approximated by the following empirical expression

$$\eta = [1 - \exp(-v)]\eta_\infty, \quad (2.88)$$

Table 2.2: Reduced screening radius, $R m_e c/\hbar$, and high-energy triplet contribution, η_∞, for electron-positron pair production obtained from the tables of Hubbell et al. (1980) as described in the text. Notice that $\hbar/m_e c = 3.8616 \times 10^{-13}$ m is the Compton wavelength of the electron.

Z	$Rm_ec/\hbar$	η_∞	Z	$Rm_ec/\hbar$	η_∞	Z	$Rm_ec/\hbar$	η_∞
1	122.81	1.157	32	33.422	1.158	63	26.911	1.194
2	73.167	1.169	33	33.068	1.157	64	26.705	1.196
3	69.228	1.219	34	32.740	1.158	65	26.516	1.197
4	67.301	1.201	35	32.438	1.158	66	26.304	1.196
5	64.696	1.189	36	32.143	1.158	67	26.108	1.197
6	61.228	1.174	37	31.884	1.166	68	25.929	1.197
7	57.524	1.176	38	31.622	1.173	69	25.730	1.198
8	54.033	1.169	39	31.438	1.174	70	25.577	1.198
9	50.787	1.163	40	31.142	1.175	71	25.403	1.200
10	47.851	1.157	41	30.950	1.170	72	25.245	1.201
11	46.373	1.174	42	30.758	1.169	73	25.100	1.202
12	45.401	1.183	43	30.561	1.172	74	24.941	1.204
13	44.503	1.186	44	30.285	1.169	75	24.790	1.205
14	43.815	1.184	45	30.097	1.168	76	24.655	1.206
15	43.074	1.180	46	29.832	1.164	77	24.506	1.208
16	42.321	1.178	47	29.581	1.167	78	24.391	1.207
17	41.586	1.175	48	29.411	1.170	79	24.262	1.208
18	40.953	1.170	49	29.247	1.172	80	24.145	1.212
19	40.524	1.180	50	29.085	1.174	81	24.039	1.215
20	40.256	1.187	51	28.930	1.175	82	23.922	1.218
21	39.756	1.184	52	28.721	1.178	83	23.813	1.221
22	39.144	1.180	53	28.580	1.179	84	23.712	1.224
23	38.462	1.177	54	28.442	1.180	85	23.621	1.227
24	37.778	1.166	55	28.312	1.187	86	23.523	1.230
25	37.174	1.169	56	28.139	1.194	87	23.430	1.237
26	36.663	1.166	57	27.973	1.197	88	23.331	1.243
27	35.986	1.164	58	27.819	1.196	89	23.238	1.247
28	35.317	1.162	59	27.675	1.194	90	23.139	1.250
29	34.688	1.154	60	27.496	1.194	91	23.048	1.251
30	34.197	1.156	61	27.285	1.194	92	22.967	1.252
31	33.786	1.157	62	27.093	1.194			

 3193

where

$$v = (0.2840 - 0.1909a)\ln(4/\kappa) + (0.1095 + 0.2206a)\ln^2(4/\kappa) + (0.02888 - 0.04269a)\ln^3(4/\kappa) + (0.002527 + 0.002623a)\ln^4(4/\kappa). \quad (2.89)$$

Then, the single quantity η_∞ characterizes the triplet production for each element.

The approximation given by eq. (2.77) with the fitted value of the screening radius, fails at low energies where it systematically underestimates the total cross section (it can even become negative). To compensate for this fact we introduce an empirical correcting term $F_0(\kappa, Z)$, which acts in a way similar to the Coulomb correction. To facilitate the random sampling, the Bethe-Heitler DCS, eq. (2.77), including this low-energy correction and a high-energy radiative correction, is written in the form

$$\frac{\mathrm{d}\sigma_{\mathrm{pp}}}{\mathrm{d}\epsilon} = r_e^2 \alpha Z[Z+\eta]\, C_r \frac{2}{3}\left[2\left(\frac{1}{2}-\epsilon\right)^2 \phi_1(\epsilon) + \phi_2(\epsilon)\right], \quad (2.90)$$

where

$$\phi_1(\epsilon) = g_1(b) + g_0(\kappa)$$
$$\phi_2(\epsilon) = g_2(b) + g_0(\kappa) \quad (2.91)$$

with

$$g_1(b) = \frac{1}{2}(3\Phi_1 - \Phi_2) - 4\ln(R m_e c/\hbar) = \frac{7}{3} - 2\ln(1+b^2) - 6b\arctan(b^{-1}) - b^2\left[4 - 4b\arctan(b^{-1}) - 3\ln(1+b^{-2})\right],$$

$$g_2(b) = \frac{1}{4}(3\Phi_1 + \Phi_2) - 4\ln(R m_e c/\hbar) = \frac{11}{6} - 2\ln(1+b^2) - 3b\arctan(b^{-1}) + \frac{1}{2}b^2\left[4 - 4b\arctan(b^{-1}) - 3\ln(1+b^{-2})\right],$$

$$g_0(\kappa) = 4\ln(R m_e c/\hbar) - 4f_C(Z) + F_0(\kappa, Z). \quad (2.92)$$

$C_r = 1.0093$ is the high-energy limit of Mork and Olsen's radiative correction (Hubbell et al., 1980).

The correcting factor $F_0(\kappa, Z)$ has been determined by requiring that the total cross section for pair production obtained from the expression given in eq. (2.90) (with $\eta = 0$) coincides with the total cross sections for pair production in the field of the nucleus tabulated by Hubbell et al. (1980). By inspection and numerical fitting, we have obtained the following analytical approximation

$$F_0(\kappa, Z) = (-0.1774 - 12.10a + 11.18a^2)(2/\kappa)^{1/2} + (8.523 + 73.26a - 44.41a^2)(2/\kappa) - (13.52 + 121.1a - 96.41a^2)(2/\kappa)^{3/2} + (8.946 + 62.05a - 63.41a^2)(2/\kappa)^2. \quad (2.93)$$

The functions ϕ_1 and ϕ_2 are now positive except for ϵ-values very near the endpoints of the allowed interval, given by eq. (2.85), for high atomic number elements. To avoid inconsistencies, these functions are set equal to zero when they take negative values.

The relative differences between the total atomic cross sections obtained from the DCS given by eq. (2.90) and the total cross sections tabulated by Hubbell et al. (1980) are appreciable near the threshold [actually, (2.90) shifts the threshold for pair production to values slightly larger than $2m_ec^2$], but decrease rapidly with increasing photon energy. At $E = 3$ MeV, the differences reduce to 4% and do not exceed 2% for energies larger than 6 MeV, for almost all the elements. Although these differences are not important, they may be larger than the uncertainties in the cross sections given by Hubbell et al. (1980). To avoid systematic errors, the mean free paths for pair production used in PENELOPE are obtained by interpolation in a table generated with the XCOM program (Berger and Hubbell, 1987). The Bethe-Heitler DCS is only used to sample the kinetic energies of the produced pair.

It is also worth noting that the Bethe-Heitler theory predicts that the pair-production DCS, considered as a function of the electron reduced energy ϵ, is symmetrical about $\epsilon = 1/2$ (see fig. 2.9). This dependence on ϵ is reasonably accurate only for photon energies larger than ~5 MeV. For lower photon energies, the effect of the electrostatic field of the atom (which slows down the electron and accelerates the positron) becomes increasingly important, with the result that the actual DCS becomes asymmetrical and the mean value of ϵ becomes less than 1/2 (see e.g. Motz et al., 1969). At these relatively low energies, however, pair production is not dominant and, morever, the produced particles have ranges that are much less than the mean free path of the absorbed photon. Therefore, no appreciable simulation errors are incurred by using the Bethe-Heitler DCS, eq. (2.90), for energies down to the threshold.

2.4.1 Simulation of pair production events

The Bethe-Heitler DCS, eq. (2.90), only depends on the kinetic energy $E_- = \epsilon E - m_ec^2$ of the produced electron, so that E_- can be directly sampled from eq. (2.90); the kinetic energy of the positron is obtained as $E_+ = E - E_- - 2m_ec^2$. Notice that, although the Bethe-Heitler total atomic cross section accounts for pair and triplet production, all the events are simulated as if they were pairs. This approximation is justified by the fact that, in triplet production, the recoiling electron has a range that is much smaller than the mean free path of the incident photon.

The electron reduced energy ϵ is distributed in the interval $(\kappa^{-1}, 1-\kappa^{-1})$, see eq. (2.85), according to the PDF given by eq. (2.90) (normalization is again irrelevant)

$$p_{\rm pp}(\epsilon) = 2\left(\frac{1}{2}-\epsilon\right)^2 \phi_1(\epsilon) + \phi_2(\epsilon), \tag{2.94}$$

which is symmetrical about the point $\epsilon = 1/2$. Fig. 2.9 shows this PDF for lead and various photon energies. The following algorithm for sampling ϵ is based on the fact

that the functions $\phi_1(\epsilon)$ and $\phi_2(\epsilon)$ are non-negative and attain their maximum values at $\epsilon = 1/2$.

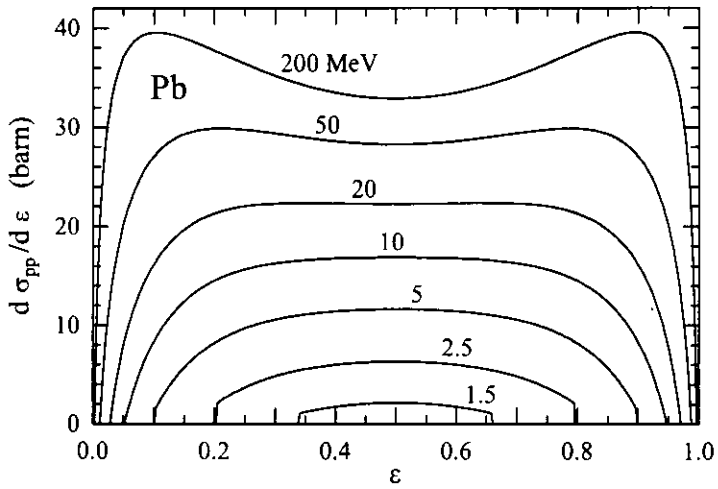


Figure 2.9: Pair production DCS in lead as a function of the electron reduced energy, $\epsilon = (E_- + m_e c^2)/E$. (Adapted from Baró et al., 1994a.)

Except for a normalization constant, the PDF (2.94) can be written in the form

$$p_{pp}(\epsilon) = u_1 U_1(\epsilon)\pi_1(\epsilon) + u_2 U_2(\epsilon)\pi_2(\epsilon) \tag{2.95}$$

with

$$u_1 = \frac{2}{3}\left(\frac{1}{2} - \frac{1}{\kappa}\right)^2 \phi_1(1/2), \qquad u_2 = \phi_2(1/2), \tag{2.96}$$

$$\pi_1(\epsilon) = \frac{3}{2}\left(\frac{1}{2} - \frac{1}{\kappa}\right)^{-3}\left(\frac{1}{2} - \epsilon\right)^2, \qquad \pi_2(\epsilon) = \frac{1}{2}\left(\frac{1}{2} - \frac{1}{\kappa}\right)^{-1} \tag{2.97}$$

and

$$U_1(\epsilon) = \phi_1(\epsilon)/\phi_1(1/2), \qquad U_2(\epsilon) = \phi_2(\epsilon)/\phi_2(1/2). \tag{2.98}$$

The functions $\pi_i(\epsilon)$ are normalized PDFs in the interval $(\kappa^{-1}, 1-\kappa^{-1})$, from which random values of ϵ can be easily sampled by using the inverse transform method. In this interval, the functions $U_i(\epsilon)$ are positive and less than unity, i.e. they are valid rejection functions. The generation of random values of ϵ from the distribution (2.95) can now be performed by combining the composition and rejection methods (see section 1.2) according to the following algorithm:

(i) Sample a value of the integer i (=1, 2) according to the point probabilities

$$p(1) = \frac{u_1}{u_1 + u_2} \qquad \text{and} \qquad p(2) = \frac{u_2}{u_1 + u_2}. \tag{2.99}$$

(ii) Sample ϵ from $\pi_i(\epsilon)$ using the sampling formulae (inverse transform method, see section 1.2.2)

$$\epsilon = \begin{cases} \frac{1}{2} + \left(\frac{1}{2} - \frac{1}{\kappa}\right)(2\xi - 1)^{1/3} & \text{if } i = 1, \\ \frac{1}{\kappa} + \left(\frac{1}{2} - \frac{1}{\kappa}\right) 2\xi & \text{if } i = 2. \end{cases} \tag{2.100}$$

(iii) Generate a new random number ξ.

(iv) If $\xi > U_i(\epsilon)$, go to step (i).

(v) Deliver ϵ.

Notice that the quantity $2\xi - 1$ may be negative and, therefore, taking its cube root will lead to a computer error; provision of this fact must be made when programming the algorithm. The efficiency of the algorithm is greater than 70% for energies near the threshold, and increases with increasing photon energies. For $E = 1$ GeV it is of the order of 95% for all the elements in the periodic table.

Angular distribution of the produced particles

Actually, the complete DCS for pair production is a function of the directions of the pair of particles. As the final state involves three bodies (the nucleus and the produced pair), the directions of the produced particles cannot be obtained from only their kinetic energies. The polar angles of the directions of movement of the electron and positron (θ_- and θ_+, fig. 2.1) relative to the direction of the incident photon are sampled from the leading term of the expression obtained from high-energy theory (Heitler, 1954; Motz et al., 1969)

$$p(\cos\theta_\pm) = a\,(1 - \beta_\pm \cos\theta_\pm)^{-2}, \tag{2.101}$$

where a is a normalization constant and

$$\beta_\pm = \frac{\sqrt{E_\pm(E_\pm + 2\mathrm{m_e}c^2)}}{E_\pm + \mathrm{m_e}c^2} \tag{2.102}$$

is the particle velocity in units of the speed of light. Random values of $\cos\theta_\pm$ are obtained by using the inverse transform method (see section 1.2.2), which leads to the sampling formula

$$\cos\theta_\pm = \frac{2\xi - 1 + \beta_\pm}{(2\xi - 1)\beta_\pm + 1}. \tag{2.103}$$

As the directions of the produced particles and the incident photon are not necessarily coplanar, the azimuthal angles ϕ_- and ϕ_+ of the electron and the positron are sampled independently and uniformly in the interval $(0, 2\pi)$.

It is worth stressing the fact that the produced charged particles have ranges that are much smaller than the mean free path of the photons. Moreover, the charged particles immediately enter a multiple elastic scattering process which randomizes their

directions of movement. As a consequence, there should be little difference between simulation results obtained with the present method and with exact random sampling from a more accurate DCS, differential in the energies and directions of the generated particles.

Compound materials

Let us consider a compound X_xY_y, in which the molecules consist of x atoms of the element X and y atoms of the element Y. The number of electrons per molecule is $Z_M = xZ(X) + yZ(Y)$ and the molecular weight is $A_M = xA_w(X) + yA_w(Y)$, where $Z(X)$ and $A_w(X)$ stand for the atomic number and atomic weight of element X.

In the simulation of pair-production events, we could use the molecular DCSs obtained from the additivity rule. The simulation of each event would then consist of 1) sampling the atom which participates in the interaction and 2) generating a random value of the electron reduced energy ϵ from the corresponding atomic DCS. To save computer time, PENELOPE generates ϵ by considering an "equivalent" single element material of the same mass density ρ as the actual medium, atomic number Z_{eq} and atomic weight A_{eq} given by

$$Z_{eq}A_M = Z_M A_{eq} = xZ(X)A_w(X) + yZ(Y)A_w(Y), \tag{2.104}$$

i.e. its atomic number (weight) is the mass-average (Z-average) of the atomic numbers (weights) of the constituent atoms. The reduced energy is sampled from the DCS of the element with the atomic number closest to Z_{eq}. Usually, this approximation does not alter the simulation results appreciably and permits a considerable simplification of the program and a reduction of the simulation time.

2.5 Attenuation coefficients

The photon inverse mean free path for a given mechanism is known as the partial attenuation coefficient of that mechanism. Thus, the partial attenuation coefficient for photoelectric absorption is

$$\mu_{ph} = \mathcal{N}\sigma_{ph}, \tag{2.105}$$

where $\mathcal{N} = N_A\rho/A_M$ is the number of atoms or molecules per unit volume and σ_{ph} is the atomic or molecular photoelectric cross section. The photoelectric mass attenuation coefficient is defined as μ_{ph}/ρ and, therefore, is independent of the density of the material. Analogous definitions apply for the other interaction processes. The total mass attenuation coefficient is obtained as

$$\frac{\mu}{\rho} = \frac{N_A}{A_M}\left(\sigma_{Ra} + \sigma_{Co} + \sigma_{ph} + \sigma_{pp}\right). \tag{2.106}$$

As mentioned above, PENELOPE uses tables of total cross sections for photoelectric absorption and pair production obtained from the database EPDL (Cullen et al., 1997)

and the program XCOM (Berger and Hubbell, 1987), respectively. Photoelectric cross sections for energies different from those in the tables are calculated by linear log-log interpolation. Total cross sections for pair production are evaluated by cubic spline log-log interpolation of the function $(1 - 2m_e c^2/E)^{-3}\sigma_{pp}$, which varies slowly with the photon energy.

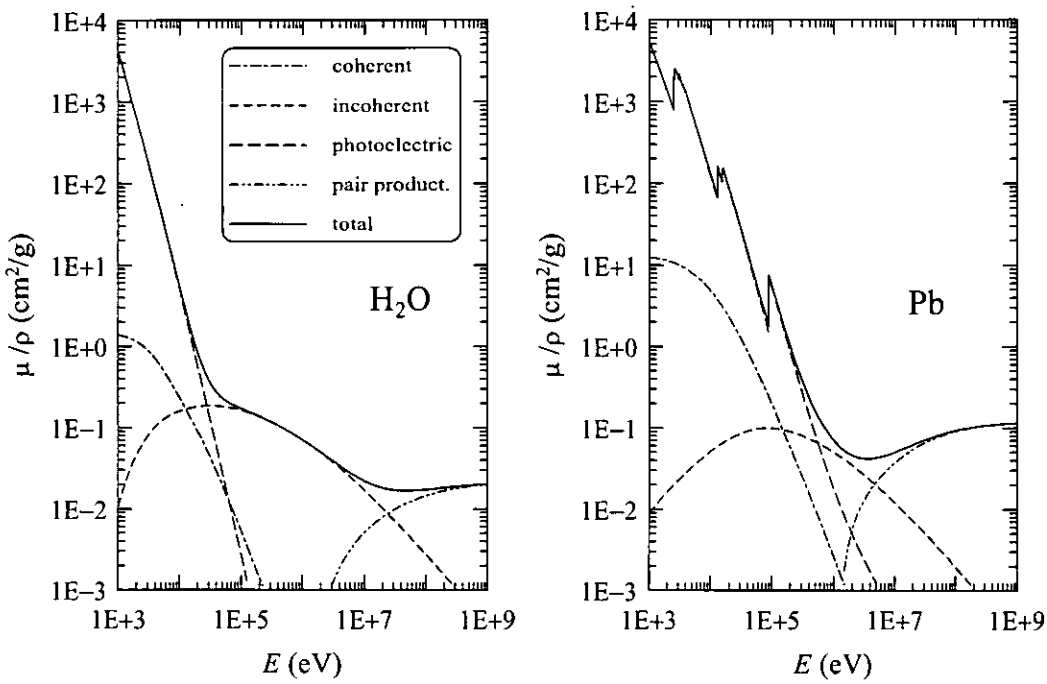


Figure 2.10: Partial and total mass attenuation coefficients of water and lead as functions of the photon energy.

Mean free paths for coherent and incoherent scattering are computed from the DCSs described in sections 2.1 and 2.3. The resulting values are virtually identical to those given by the XCOM program for E greater than $\sim$ 50 keV. At lower energies, our mean free paths for Compton scattering deviate from those given by XCOM; these were calculated from a different theoretical model (Hubbell et al., 1975), which neglects Doppler broadening (see e.g. Brusa et al., 1996). The evaluation of the total atomic cross section for these processes [see eqs. (2.10) and (2.54)] involves a numerical quadrature, which is performed by using the function SUMGA (appendix B). Notice that for high-energy photons, the integrand in the coherent scattering cross section, eq. (2.10), is sharply peaked at $\theta = 0$. In such a case, the numerical integration method is not effective. For energies larger than $\sim Z/2$ MeV, we take advantage of the asymptotic behaviour shown by eq. (2.12) to avoid time-consuming integration. Partial and total attenuation coefficients for water and lead, as representatives of low- and high-Z materials, are displayed in fig. 2.10.

2.6 Atomic relaxation

Atoms are primarily ionized by photon interactions and by electron or positron impact. There is a fundamental difference between the ionizing effects of photons and of charged particles. A photon is only able to directly ionize a few atoms. In the case of photoabsorption, when the photon energy is larger than the K-shell binding energy, about 80% of photoabsorptions occur in the K-shell, i.e. the resulting ion with a vacancy in the K-shell is highly excited. Incoherent scattering is not as highly preferential, but still the probability that an inner shell is ionized is nearly proportional to the number of electrons in the shell. Conversely, fast electrons and positrons (and other charged particles) ionize many atoms along their paths; the ionizations occur preferentially in the less tightly bound atomic shells, or the conduction band in the case of metals (see section 3.2), so that most of the produced ions are only weakly excited.

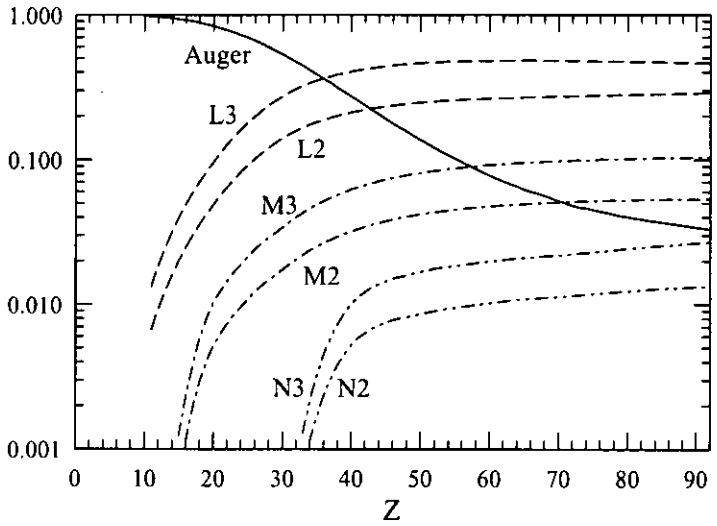


Figure 2.11: Relative probabilities for radiative and non-radiative (Auger) transitions that fill a vacancy in the K-shell of atoms.

Excited ions with a vacancy in an inner shell relax to their ground state through a sequence of radiative and non-radiative transitions. In a radiative transition, the vacancy is filled by an electron from an outer shell and an x-ray with characteristic energy is emitted. In a non-radiative transition, the vacancy is filled by an outer electron and the excess energy is released through emission of an electron from a shell that is farther out (Auger effect). Each non-radiative transition generates an additional vacancy that, in turn, migrates "outwards". The production of vacancies in inner shells and their subsequent relaxation must be simulated in detail, since the energetic x-rays and/or electrons emitted during the process may transport energy quite a distance from the excited ion.

PENELOPE simulates the emission of characteristic radiation and Auger electrons that

result from vacancies produced in K-shells and L-subshells by photoelectric absorption, Compton scattering and electron/positron impact (see chapter 3). The relaxation of these vacancies is followed until the K- and L-shells are filled up, i.e. until the vacancies have migrated to M and outer shells. Vacancies in these outer shells originate much less energetic secondary radiation, whose main effect is to spread out the excitation energy of the ion within the surrounding material. To get a reliable description of the dose distribution, and other macroscopic transport characteristics, we only have to follow secondary radiation that is able to propagate to distances of the order of, say, 1% of the penetration distance (or range) of the primary radiation. Radiation with lower energy does not need to be followed, since its only effect is to blur the "primary" dose distribution on a small length scale.

To simplify the description of the ionization processes of outer shells (i.e. photoelectric absorption, Compton scattering and electron/positron impact), we simply assume that, when ionization occurs in M or outer shells, a secondary (delta) electron is emitted from the parent ion with a kinetic energy E_s equal to the energy deposited by the primary particle,

$$E_{\text{dep}} = \begin{cases} E - E' & \text{in Compton scattering,} \\ E & \text{in photoelectric absorption,} \\ W & \text{in electron/positron impact (see chapter 3).} \end{cases} \tag{2.107}$$

That is, the whole excitation energy of the ion is taken up by the ejected electron and no fluorescent radiation is simulated. In reality, the emitted electrons have energies less than the values (2.107) and can be followed by characteristic x-rays, which have mean free paths that are usually much larger than the Bethe range of photoelectrons. By giving an artificially increased initial energy to the electron we allow it to transport energy farther from the ion so as to partially compensate for the neglect of other radiation emitted during the de-excitation cascade.

In the case of ionization of an inner shell i, i.e. a K-shell or an L-shell, we consider that the electron is ejected with kinetic energy

$$E_s = E_{\text{dep}} - U_i, \tag{2.108}$$

where U_i is the ionization energy of the active shell, and that the target atom is left with a vacancy in shell i. As mentioned above, we consider only characteristic x-rays and Auger electrons emitted in the first stages of the relaxation process. These secondary radiations are assumed to be emitted isotropically from the excited atom. We use the following notation to designate the possible transitions

- Radiative: S0-S1 (an electron from the S1 shell fills the vacancy in the S0 shell, leaving a hole in the S1 shell). The considered radiative transitions (for elements with $Z > 18$ with the M-shell filled) are shown in fig. 2.3.
- Non-radiative: S0-S1-S2 (an electron from the S1 shell fills the vacancy in the S0 shell, and the released energy is taken away by an electron in the S2 shell; this process leaves two vacancies, in the S1 and S2 shells).

Non-radiative transitions of the type LI-LJ-Xq, which involve an electron transition between two L-subshells and the ejection of an electron from an outer shell Xq are known as L-shell Coster-Kronig transitions. .

The information furnished to PENELOPE for each element consists of a table of possible transitions, transition probabilities and energies of the emitted x-rays or electrons for ionized atoms with a single vacancy in the K-shell or in an L-subshell. These data are entered through the material definition file. The transition probabilities are extracted from the LLNL Evaluated Atomic Data Library (Perkins et al., 1991). Fig. 2.11 displays transition probabilities for the transitions that fill a vacancy in the K shell as functions of the atomic number Z; the curve labelled "Auger" corresponds to the totality of non-radiative transitions. We see that for low-Z elements, the relaxation proceeds mostly through non-radiative transitions. It is worth noting that the ratio of probabilities of the radiative transitions K-S2 and K-S3 (where S stands for L, M or N) is approximately 1/2, as obtained from the dipole approximation (see e.g. Bransden and Joachain, 1983); radiative transitions K-S1 are strictly forbidden (to first order) within the dipole approximation.

The energy of the x-ray emitted in the radiative transition S0-S1 is assumed to be

$$E_{\mathrm{x}} = U_{\mathrm{S0}} - U_{\mathrm{S1}}, \tag{2.109}$$

where $U_{\mathrm{S}i}$ is the binding energy of an electron in the shell Si of the neutral atom, which is taken from the PENELOPE database. Similarly, the energy of the electron emitted in the non-radiative transition S0-S1-S2 is set equal to

$$E_{\mathrm{e}} = U_{\mathrm{S0}} - U_{\mathrm{S1}} - U_{\mathrm{S2}}. \tag{2.110}$$

These emission energies correspond to assuming that the presence of the vacancy (or vacancies) does not alter the ionization energies of the active electron shells, which is an approximation. It should be noted that these formulae are also used to determine the energies of the emitted radiation at any stage of the de-excitation cascade, which means that we neglect the possible relaxation of the ion (see e.g. Sevier, 1972). Therefore, our approach will not produce L_α and L_β x-ray satellite lines; these arise from the filling of a vacancy in a doubly-ionized L-shell (generated e.g. by a Coster-Kronig transition), which releases an energy that is slightly different from the energy liberated when the shell contains only a single vacancy. It is also worth recalling that the adopted transition probabilities are approximate. For K shells they are expected to be accurate to within one per cent or so, but for other shells they are subject to much larger uncertainties. Even the L-shell fluorescence yield (the sum of radiative transition probabilities for an L-shell vacancy) is uncertain by about 20% (see e.g. Hubbell, 1989; Perkins et al., 1991).

The simulation of the relaxation cascade is performed by subroutine **RELAX**. The transition that fills the initial vacancy is randomly selected according to the adopted transition probabilities, by using Walker's aliasing method (section 1.2.3). This transition leaves the ion with one or two vacancies. If the energy of the emitted characteristic x-ray or Auger electron is larger than the corresponding absorption energy, the state

variables of the particle are stored in the secondary stack (which contains the initial states of all particles produced during the current shower that have not yet been simulated). The generation of the cascade continues by repeating the process for each remaining vacancy. It ends either when the K-shell and L-subshells have been filled up or when there is not enough energy to produce "active" radiation (with energy larger than the absorption energy). The excitation energy of the residual ion is assumed to be deposited locally.

It is important to bear in mind that we are disregarding the emission and transport of soft x-rays and slow electrons. This sets a lower limit to the photon energies for which PENELOPE is applicable. In principle, simulation results are expected to be reliable only for photons with energies larger than the ionization energy of the M1 subshell of the heaviest element present (125 eV for copper, 720 eV for silver, 3.4 keV for gold and 5.5 keV for uranium).

Chapter 3

Electron and positron interactions

In this chapter we consider the interactions of fast electrons and positrons of kinetic energy E with matter. For the sake of simplicity, we start by assuming that the particles move in a single-element medium of atomic number Z and density ρ, with $\mathcal{N}$ atoms per unit volume. The extension to compounds, and mixtures, is normally done on the basis of the additivity approximation, i.e. the molecular DCS is approximated as the *incoherent* sum of the atomic DCSs of all the atoms in a molecule.

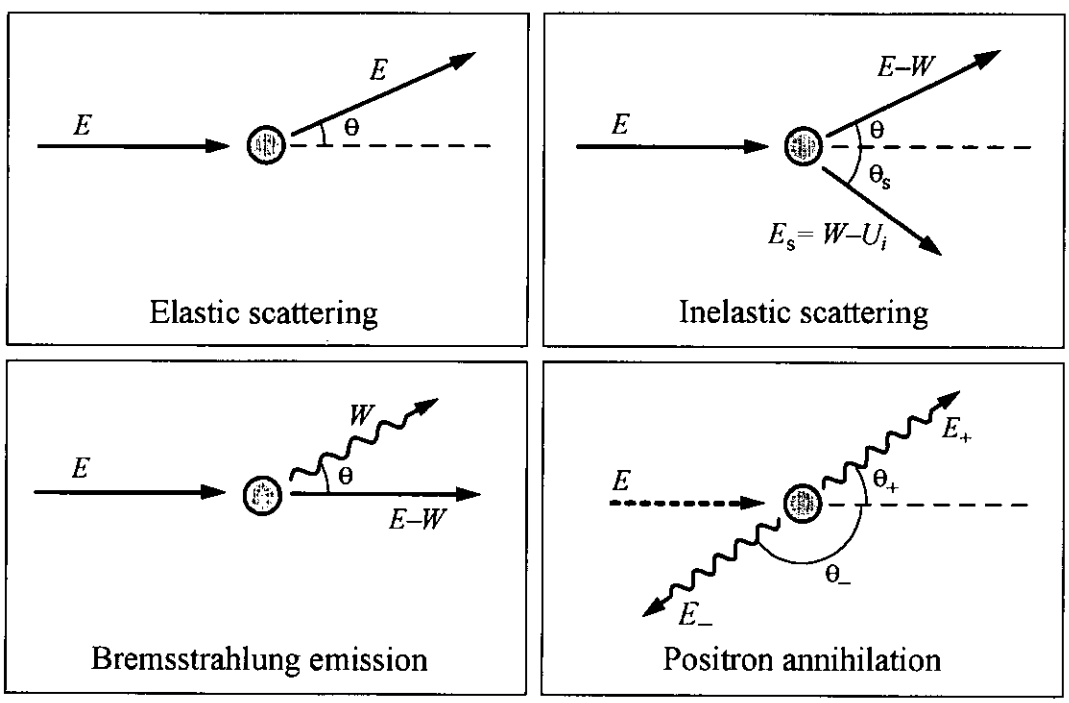


Figure 3.1: Basic interactions of electrons and positrons with matter.

The possible interactions of electrons and positrons with the medium are elastic scattering, inelastic collisions and bremsstrahlung emission; positrons can also undergo annihilation, either in flight or at rest. The atomic DCSs adopted in PENELOPE are defined either as analytical functions or by means of numerical tables, or as a combination of both. These DCSs, which are sufficiently accurate for most practical simulation purposes, allow fast and accurate random sampling of the individual interactions. It is worth pointing out that multiple scattering distributions are quite insensitive to the fine details of the single scattering DCSs. If the adopted DCSs have a physically reasonable shape, only the values of a few integrals of the DCS have a direct influence on the simulation results (Liljequist, 1987; Fernández-Varea et al., 1993b). As a consequence, a general-purpose simulation procedure can be made fairly simple by using approximate DCSs with the proviso that they exactly reproduce the correct values of the relevant integrals. The DCSs described below represent a compromise between reliability and simplicity; they are simple enough to allow the use of fast sampling methods and, at the same time, they are flexible enough to account for the relevant features of the interactions.

Owing to the large number of interactions suffered by a fast electron or positron before coming to rest, detailed simulation is unfeasible at high energies. In PENELOPE we overcome this practical difficulty by using a mixed simulation procedure (see chapter 4) instead of the habitual condensed simulation schemes adopted in other high-energy simulation codes —e.g. ETRAN (Berger and Seltzer, 1988), ITS3 (Halbleib et al., 1992), EGS4 (Nelson et al., 1985), GEANT (Brun et al., 1986). The formulation of mixed simulation is complicated by the fact that the sampling of hard interactions is done from restricted DCSs, with cutoffs that vary with the particle energy during the evolution of a track. This limits the complexity of the DCSs that can be efficiently used in a simulation code.

3.1 Elastic collisions

In this section we consider the theoretical description of elastic collisions of electrons and positrons with isolated neutral atoms of atomic number Z at rest. By definition, elastic interactions are those in which the initial and final quantum states of the target atom are the same, normally the ground state. The angular deflections of electron trajectories in matter are mainly (but not completely) due to elastic scattering. Notice that there is a certain energy transfer from the projectile to the target, which causes the recoil of the latter (see section A.1.1). Because of the large mass of the target ($\sim 3500 Z m_e$), the average energy lost by the projectile is a very small fraction of its initial energy (a few meV for scattering of 30 keV electron by Al atoms) and is usually neglected, which is equivalent to assuming that the target has an infinite mass and does not recoil.

For a wide energy range (say from a few hundred eV to ~1 GeV), elastic interactions can be described as scattering of the projectile by the electrostatic field of the target (Mott and Massey, 1965). The charge distribution of the target atom consists of the

nucleus and the electron cloud. The density of atomic electrons $\rho(\mathbf{r})$ can be calculated by using available Hartree-Fock codes (e.g. the one of Desclaux, 1975). For atoms with closed shell configurations, the electron distribution is spherically symmetrical; for atoms with open shells, we assume that an average over directions is performed to give a spherical density $\rho(r)$. To account for the effect of the finite size of the nucleus on the elastic DCS (which is appreciable only for projectiles with energy E larger than a few MeV), we can represent the nucleus as a uniformly charged sphere of radius

$$R_{\text{nuc}} = 1.05 \times 10^{-15} A_{\text{w}}^{1/3} \text{ m}, \tag{3.1}$$

where A_{w} is the atomic mass (in g/mol). The electrostatic potential of the target atom is

$$\varphi(r) = \varphi_{\text{nuc}}(r) - e\,4\pi \left[\frac{1}{r}\int_0^r \rho(r')r'^2\,\mathrm{d}r' + \int_r^\infty \rho(r')r'\,\mathrm{d}r'\right], \tag{3.2}$$

where

$$\varphi_{\text{nuc}}(r) = \begin{cases} \dfrac{1}{2}\dfrac{Ze}{R_{\text{nuc}}}\left[3 - \left(\dfrac{r}{R_{\text{nuc}}}\right)^2\right] & \text{if } r \le R_{\text{nuc}}, \\ \dfrac{Ze}{r} & \text{if } r > R_{\text{nuc}} \end{cases} \tag{3.3}$$

is the potential of the nucleus.

Within the static-field approximation (Mott and Massey, 1965; Walker, 1971), the DCS for elastic scattering of electrons or positrons is obtained by solving the partial-wave expanded Dirac equation for the motion of the projectile in the field of the target atom. The interaction energy is given by

$$V(r) = z_0 e\,\varphi(r) + V_{\text{ex}}(r), \tag{3.4}$$

where z_0 is the charge of the projectile in units of e (-1 for electrons, $+1$ for positrons). The term $V_{\text{ex}}(r)$, which applies only for electrons, represents a local approximation to the exchange interaction between the projectile and the atomic electrons (see e.g. Salvat, 1998). We shall limit our considerations to the case of spin unpolarized projectiles, i.e. their spin is randomly oriented. Then, the effect of elastic interactions can be described as a deflection of the projectile trajectory, characterized by the polar and azimuthal scattering angles θ and ϕ. For a central field, the angular distribution of singly scattered electrons is axially symmetric about the direction of incidence, i.e. independent of ϕ. The DCS (per unit solid angle) for elastic scattering of a projectile with kinetic energy E into the solid angle element $\mathrm{d}\Omega$ about the direction (θ, ϕ) is given by (Walker, 1971)

$$\frac{\mathrm{d}\sigma_{\text{el}}}{\mathrm{d}\Omega} = |f(\theta)|^2 + |g(\theta)|^2, \tag{3.5}$$

where

$$\begin{aligned} f(\theta) &= \frac{1}{2\mathrm{i}k}\sum_{\ell=0}^{\infty}\left\{(\ell+1)\left[\exp(2\mathrm{i}\delta_{\ell+}) - 1\right] + \ell\left[\exp(2\mathrm{i}\delta_{\ell-}) - 1\right]\right\} P_\ell(\cos\theta), \\ g(\theta) &= \frac{1}{2\mathrm{i}k}\sum_{\ell=0}^{\infty}\left\{\exp(2\mathrm{i}\delta_{\ell-}) - \exp(2\mathrm{i}\delta_{\ell+})\right\} P_\ell^1(\cos\theta) \end{aligned} \tag{3.6}$$

are the direct and spin-flip scattering amplitudes, respectively.

$$k \equiv \frac{p}{\hbar} = \frac{1}{\hbar c}\left[E(E + 2m_e c^2)\right]^{1/2} \tag{3.7}$$

is the wave number of the projectile, $P_\ell(\cos\theta)$ are Legendre polynomials, $P_\ell^1(\cos\theta)$ are associated Legendre functions and $\delta_{\ell\pm}$ are the phase shifts. These are determined from the asymptotic behaviour of the Dirac radial functions for large r (Walker, 1971). Thus, to determine each phase shift we must solve the radial Dirac equations for the potential $V(r)$. The convergence of the partial-wave series (3.6) slows down when the energy of the projectile increases. This makes the calculation difficult for energies larger than a few MeV (in the case of scattering by gold atoms, about 10,000 phase shifts are required at $E = 10$ MeV). The partial-wave DCS, eq. (3.5), rigourously accounts for spin and other relativistic effects, as well as finite nuclear size effects. The elastic scattering database used in PENELOPE has been calculated essentially by this method, using a computer code written by Salvat (2000).

Single elastic collisions are determined by the values of the polar and azimuthal scattering angles, θ and ϕ, respectively. Owing to the assumed spherical symmetry of the scattering centres, single and multiple scattering angular distributions are axially symmetrical about the direction of incidence, i.e. they are independent of the azimuthal scattering angle ϕ. For simulation purposes, it is convenient to measure polar angular deflections produced by single scattering events in terms of the variable [see eq. (1.61)]

$$\mu = (1 - \cos\theta)/2 \tag{3.8}$$

instead of the scattering angle θ. Notice that μ varies from 0 (forward scattering) to 1 (backward scattering). The DCS per unit angular deflection is

$$\frac{d\sigma_{el}}{d\mu} = 4\pi \frac{d\sigma_{el}}{d\Omega}. \tag{3.9}$$

Fig. 3.2 displays DCSs for elastic scattering of electrons and positrons of various energies by aluminium and gold atoms. These numerical results illustrate the variation of the DCS with the atomic number Z, the charge of the projectile and the energy E. Since the interaction $V(r)$ is attractive for electrons and repulsive for positrons, the scattering is more intense for electrons (which can fall deeply into the potential well of the atom) than for positrons (which are repelled from the nucleus and cannot "see" the inner part of the atom). The DCS for low-energy electrons exhibits a diffraction-like structure, while the DCS for positrons decreases monotonously with the deflection μ. The Born approximation (see e.g. Mott and Massey, 1965) predicts a structureless DCS that decreases with μ and is proportional to the squared charge of the projectile (i.e. the same for electrons and positrons). This approximation considers the scattering field as a perturbation (to first order) and, hence, it is valid only for weak fields (low-Z elements). The difference between the (partial wave) DCSs for electrons and positrons gives a clear indication of the applicability of the Born approximation.

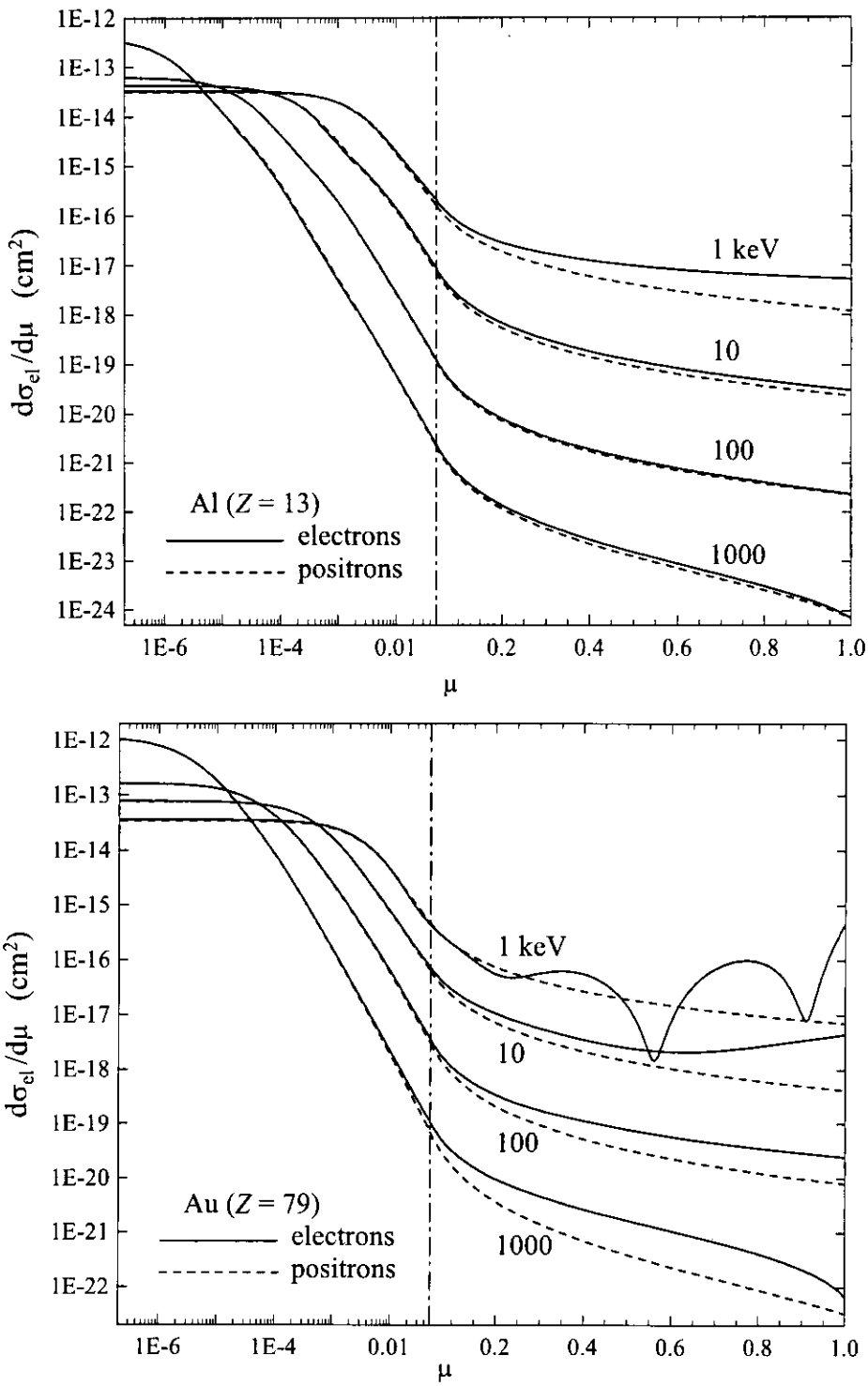


Figure 3.2: DCS for elastic scattering of electrons and positrons by aluminium and gold atoms as a function of the deflection $\mu = (1 - \cos\theta)/2$. Notice the change from logarithmic to linear scale at $\mu = 0.05$.

The total elastic cross section is given by

$$\sigma_{\rm el} = \int \frac{{\rm d}\sigma_{\rm el}}{{\rm d}\Omega}\,{\rm d}\Omega = \int \frac{{\rm d}\sigma_{\rm el}}{{\rm d}\mu}\,{\rm d}\mu. \tag{3.10}$$

Notice that we can write

$$\frac{{\rm d}\sigma_{\rm el}}{{\rm d}\mu} = \sigma_{\rm el}\, p_{\rm el}(\mu), \tag{3.11}$$

where $p_{\rm el}(\mu)$ is the normalized PDF of μ in a single collision. The mean free path between consecutive elastic events in a homogeneous single-element medium is

$$\lambda_{\rm el} = 1/(\mathcal{N}\sigma_{\rm el}), \tag{3.12}$$

where $\mathcal{N}$ is the number of atoms per unit volume.

Other important quantities (see section 4.1) are the transport cross sections[1]

$$\sigma_{{\rm el},\ell} \equiv \int \left[1 - P_\ell(\cos\theta)\right] \frac{{\rm d}\sigma_{\rm el}}{{\rm d}\Omega}\,{\rm d}\Omega. \tag{3.13}$$

The ℓ-th transport mean free path is defined by

$$\lambda_{{\rm el},\ell} \equiv 1/(\mathcal{N}\sigma_{{\rm el},\ell}). \tag{3.14}$$

The first and second transport cross sections, $\sigma_{{\rm el},1}$ and $\sigma_{{\rm el},2}$, are given by

$$\sigma_{{\rm el},1} = \int (1-\cos\theta)\frac{{\rm d}\sigma_{\rm el}}{{\rm d}\Omega}\,{\rm d}\Omega = 2\sigma_{\rm el}\int_0^1 \mu p_{\rm el}(\mu)\,{\rm d}\mu = 2\sigma_{\rm el}\,\langle\mu\rangle \tag{3.15}$$

and

$$\sigma_{{\rm el},2} = \int \frac{3}{2}(1-\cos^2\theta)\frac{{\rm d}\sigma_{\rm el}}{{\rm d}\Omega}\,{\rm d}\Omega = 6\sigma_{\rm el}\int_0^1 (\mu-\mu^2)p_{\rm el}(\mu)\,{\rm d}\mu = 6\sigma_{\rm el}\left(\langle\mu\rangle - \langle\mu^2\rangle\right), \tag{3.16}$$

where $\langle\cdots\rangle$ indicates the average value in a single collision. The quantities $\lambda_{{\rm el},1}$ and $\lambda_{{\rm el},2}$, eq. (3.14), determine the first and second moments of the multiple scattering distributions (see section 4.1). The inverse of the first transport mean free path,

$$\lambda_{{\rm el},1}^{-1} = \mathcal{N}\sigma_{{\rm el},1} = \frac{2}{\lambda_{\rm el}}\langle\mu\rangle, \tag{3.17}$$

gives a measure of the average angular deflection per unit path length. By analogy with the "stopping power", which is defined as the mean energy loss per unit path length (see section 3.2.3), the quantity $2\lambda_{{\rm el},1}^{-1}$ is sometimes called the "scattering power"[2].

[1]The Legendre polynomials of lowest orders are

$$P_0(x) = 1, \quad P_1(x) = x, \quad P_2(x) = \frac{1}{2}(3x^2-1).$$

[2]At high energies, where the scattering is concentrated at very small angles, $\langle\mu\rangle \simeq \langle\theta^2\rangle/4$ and $\lambda_{{\rm el},1}^{-1} \simeq \langle\theta^2\rangle/(2\lambda_{\rm el})$.

Fig. 3.3 shows elastic mean free paths and transport mean free paths for electrons in aluminium and gold. At low energies, the differences between the DCS of the two elements (see fig. 3.2) produce very visible differences between the transport mean free paths. When E increases, the DCS becomes strongly peaked in the forward direction and $\langle\mu^2\rangle$ becomes much smaller than $\langle\mu\rangle$. In the high-energy limit, $\sigma_{el,2} \simeq 3\sigma_{el,1}$ ($\lambda_{el,2} \simeq \lambda_{el,1}/3$). The total cross section, $\propto 1/(\rho\lambda_{el})$, decreases monotonously with E to reach a constant value at high energies. This saturation is a relativistic effect: the total cross section measures the interaction probability, which is proportional to the time spent by the projectile within the region where the scattering field is appreciable. This time is determined by the speed of the projectile, which approaches c from below when the projectile energy increases. In the non-relativistic theory, the speed $v_{n.r.} = (2E/m_e)^{1/2}$ increases without limit with E and the calculated non-relativistic total cross section tends to zero at high energies.

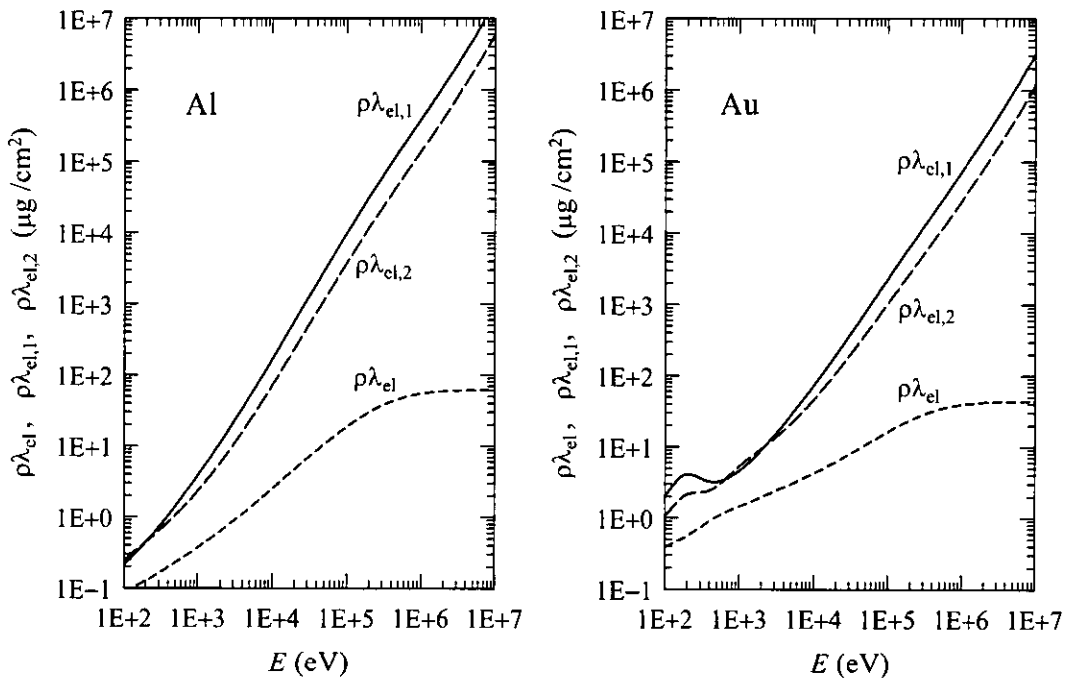


Figure 3.3: Elastic mean free path, λ_{el}, and first and second transport mean free paths, $\lambda_{el,1}$ and $\lambda_{el,2}$, for electrons scattered in aluminium and gold as functions of the kinetic energy of the projectile.

3.1.1 The modified Wentzel (MW) model

Although it is possible to do Monte Carlo simulation of electron and positron transport using numerical partial-wave DCSs (Benedito et al., 2001), this procedure is too

laborious to be adopted as the basis of a simulation code for general purposes (mostly because of the large volume of required numerical information). It is more convenient to use suitable analytical approximate DCSs that may differ in detail from the partial wave DCSs but lead to nearly the same multiple scattering distributions. In PENELOPE we use a model in which the DCS is expressed as

$$\frac{\mathrm{d}\sigma_{\mathrm{el}}^{(\mathrm{MW})}}{\mathrm{d}\mu} = \sigma_{\mathrm{el}}\, p_{\mathrm{MW}}(\mu). \tag{3.18}$$

The single scattering distribution $p_{\mathrm{MW}}(\mu)$ is defined by a simple analytical expression, with a physically plausible form, depending on two adjustable parameters. These parameters are determined in such a way that the values of $\langle\mu\rangle$ and $\langle\mu^2\rangle$ obtained from $p_{\mathrm{MW}}(\mu)$ are equal to those of the actual (partial wave) DCS:

$$\langle\mu\rangle_{\mathrm{MW}} \equiv \int_0^1 \mu p_{\mathrm{MW}}(\mu)\,\mathrm{d}\mu = \langle\mu\rangle = \frac{1}{2}\frac{\sigma_{\mathrm{el},1}}{\sigma_{\mathrm{el}}} \tag{3.19}$$

and

$$\langle\mu^2\rangle_{\mathrm{MW}} \equiv \int_0^1 \mu^2 p_{\mathrm{MW}}(\mu)\,\mathrm{d}\mu = \langle\mu^2\rangle = \frac{1}{2}\frac{\sigma_{\mathrm{el},1}}{\sigma_{\mathrm{el}}} - \frac{1}{6}\frac{\sigma_{\mathrm{el},2}}{\sigma_{\mathrm{el}}}. \tag{3.20}$$

Thus, the MW model will give the same mean free path and the same first and second transport mean free paths as the partial wave DCS. As a consequence (see chapter 4), detailed simulations using this model will yield multiple scattering distributions that do not differ significantly from those obtained from the partial wave DCS, quite irrespectively of other details of the "artificial" distribution $p_{\mathrm{MW}}(\mu)$.

To set the distribution $p_{\mathrm{MW}}(\mu)$, we start from the Wentzel (1927) angular distribution,

$$p_{\mathrm{W},A_0}(\mu) \equiv \frac{A_0(1+A_0)}{(\mu+A_0)^2}, \qquad (A_0 > 0) \tag{3.21}$$

which describes the scattering by an exponentially screened Coulomb field within the Born approximation (see e.g. Mott and Massey, 1965), that is, it provides a physically plausible angular distribution, at least for light elements or high-energy projectiles. It is also worth mentioning that the multiple scattering theory of Molière (1947, 1948) can be derived by assuming that electrons scatter according to the Wentzel distribution (see Fernández-Varea et al., 1993b). The first moments of the Wentzel distribution are

$$\langle\mu\rangle_{\mathrm{W},A_0} = \int_0^1 \mu\,\frac{A_0(1+A_0)}{(\mu+A_0)^2}\,\mathrm{d}\mu = A_0\left[(1+A_0)\ln\left(\frac{1+A_0}{A_0}\right) - 1\right] \tag{3.22}$$

and

$$\langle\mu^2\rangle_{\mathrm{W},A_0} = \int_0^1 \mu^2\,\frac{A_0(1+A_0)}{(\mu+A_0)^2}\,\mathrm{d}\mu = A_0\left[1 - 2\langle\mu\rangle_{\mathrm{W},A_0}\right]. \tag{3.23}$$

Let us define the value of the screening constant A_0 so that $\langle\mu\rangle_{\mathrm{W},A_0} = \langle\mu\rangle$. The value of A_0 can be easily calculated by solving eq. (3.22) numerically, e.g. by the Newton-Raphson method. Usually, we shall have $\langle\mu^2\rangle_{\mathrm{W},A_0} \neq \langle\mu^2\rangle$. At low energies, the Wentzel

distribution that gives the correct average deflection is too "narrow" [$\langle\mu^2\rangle_{\mathrm{W},A_0} < \langle\mu^2\rangle$ for both electrons and positrons and for all the elements]. At high energies, the angular distribution is strongly peaked in the forward direction and the Wentzel distribution becomes too "wide". This suggests using a modified Wentzel (MW) model obtained by combining a Wentzel distribution with a simple distribution, which takes different forms in these two cases,

• Case I. If $\langle\mu^2\rangle_{\mathrm{W},A_0} > \langle\mu^2\rangle$ (the Wentzel distribution is too wide), we take $p_{\mathrm{MW}}(\mu)$ as a statistical admixture of the Wentzel distribution and a delta distribution (a zero-width, fixed scattering angle process)

$$p_{\mathrm{MW,I}}(\mu) = (1-B)\,p_{\mathrm{W},A}(\mu) + B\,\delta(\mu - \langle\mu\rangle) \tag{3.24}$$

with

$$A = A_0 \qquad \text{and} \qquad B = \frac{\langle\mu^2\rangle_{\mathrm{W},A} - \langle\mu^2\rangle}{\langle\mu^2\rangle_{\mathrm{W},A} - \langle\mu\rangle^2}. \tag{3.25}$$

Notice that in this case we usually have $\langle\mu\rangle \ll 1$, so that the delta distribution is at very small angles. Although we have introduced a discrete peak in the DCS, its effect is smeared out by the successive collisions and not visible in the multiple scattering angular distributions.

• Case II. If $\langle\mu^2\rangle_{\mathrm{W},A_0} < \langle\mu^2\rangle$ (the Wentzel distribution is too narrow), we express $p_{\mathrm{MW}}(\mu)$ as a statistical admixture of a Wentzel distribution (with A not necessarily equal to A_0) and a triangle distribution in the interval (1/2,1),

$$p_{\mathrm{MW,II}}(\mu) = (1-B)\,p_{\mathrm{W},A}(\mu) + B\,8\,(\mu - 1/2)\,\Theta\,(\mu - 1/2)\,. \tag{3.26}$$

The parameters A and B are obtained from the conditions (3.19) and (3.20), which give

$$\begin{aligned}(1-B)\,\langle\mu\rangle_{\mathrm{W},A} + B\,\frac{5}{6} &= \langle\mu\rangle \\ (1-B)\,\langle\mu^2\rangle_{\mathrm{W},A} + B\,\frac{17}{24} &= \langle\mu^2\rangle.\end{aligned} \tag{3.27}$$

From the first of these equations,

$$B = \frac{\langle\mu\rangle - \langle\mu\rangle_{\mathrm{W},A}}{(5/6) - \langle\mu\rangle_{\mathrm{W},A}}. \tag{3.28}$$

Inserting this value in the second of eqs. (3.27), we obtain

$$\left(\frac{17}{24} - \langle\mu^2\rangle\right)\langle\mu\rangle_{\mathrm{W},A} - \left(\frac{5}{6} - \langle\mu\rangle\right)\langle\mu^2\rangle_{\mathrm{W},A} = \frac{17}{24}\langle\mu\rangle - \frac{5}{6}\langle\mu^2\rangle. \tag{3.29}$$

For all situations of interest, this equation has a single root A in the interval $(0, A_0)$ and can be easily solved by means of a bipartition procedure. The value of B given by eq. (3.28) is then positive and less than unity, as required.

In fig. 3.4 we compare partial wave DCSs and MW model DCSs for elastic scattering of electrons of various energies by gold atoms. The considered energies correspond to

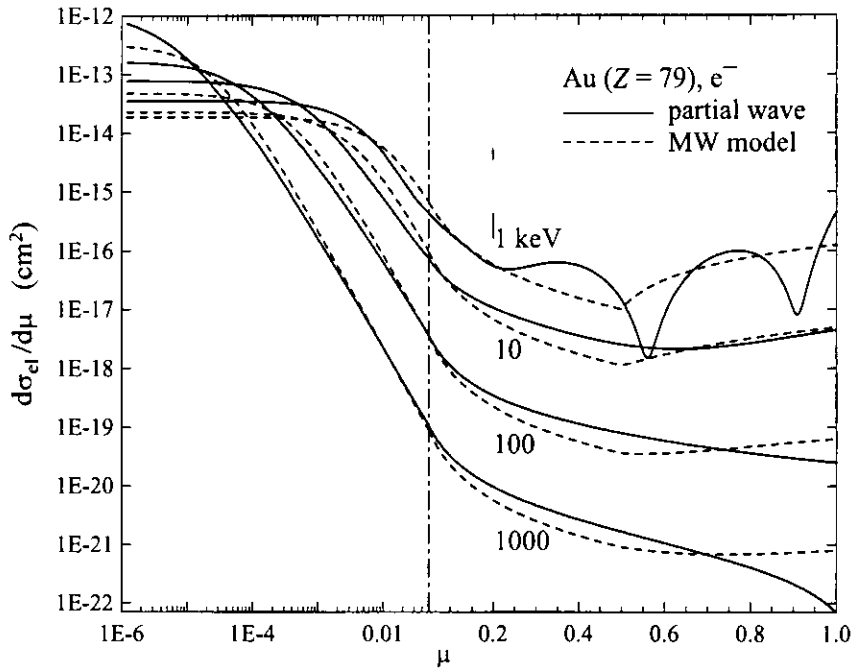


Figure 3.4: Partial wave and MW model DCSs for elastic scattering of electrons by gold atoms.

the case-II MW model [so that the distribution $p_{\rm MW}(\mu)$ is continuous]. We see that the MW model does imitate the partial wave DCSs, but the differences are significant. Nevertheless, the important fact here is that both DCSs give exactly the same values of $\sigma_{\rm el}$, $\langle\mu\rangle$ and $\langle\mu^2\rangle$.

The information needed to determine the parameters of the MW model reduces to the characteristic functions $\sigma_{\rm el}(E)$, $\sigma_{{\rm el},1}(E)$ and $\sigma_{{\rm el},2}(E)$. PENELOPE reads these functions from a precalculated database for electrons and positrons, for the elements Z = 1–92 and for a grid of energies that is dense enough to permit accurate cubic spline log-log interpolation. This elastic scattering database was generated by using the partial-wave code of Salvat (2000); the atomic electron densities were obtained from the Dirac-Hartree-Fock code of Desclaux (1975) —which correspond to free atoms. Before starting the simulation, PENELOPE evaluates a table of the parameters A and B, and stores it in the computer memory. Instead of B, PENELOPE tabulates the quantity $B' = +B$ (case I) and $B' = -B$ (case II); this avoids the need to specify the case, which can be inferred from the sign of B'. It is worth noting that A and B' are continuous functions of energy and, therefore, can be rapidly evaluated, for any energy, by interpolation in the stored table. In case I, $\langle\mu\rangle$ concides with $\langle\mu\rangle_{{\rm W},A}$, which is determined by A, eq. (3.22). Fig. 3.5 displays the MW model parameters for aluminium and gold, as representative of low- and high-Z elements. Notice that at high energies, where the case I model applies, the strength of the delta contribution increases rapidly with energy, indicating that the

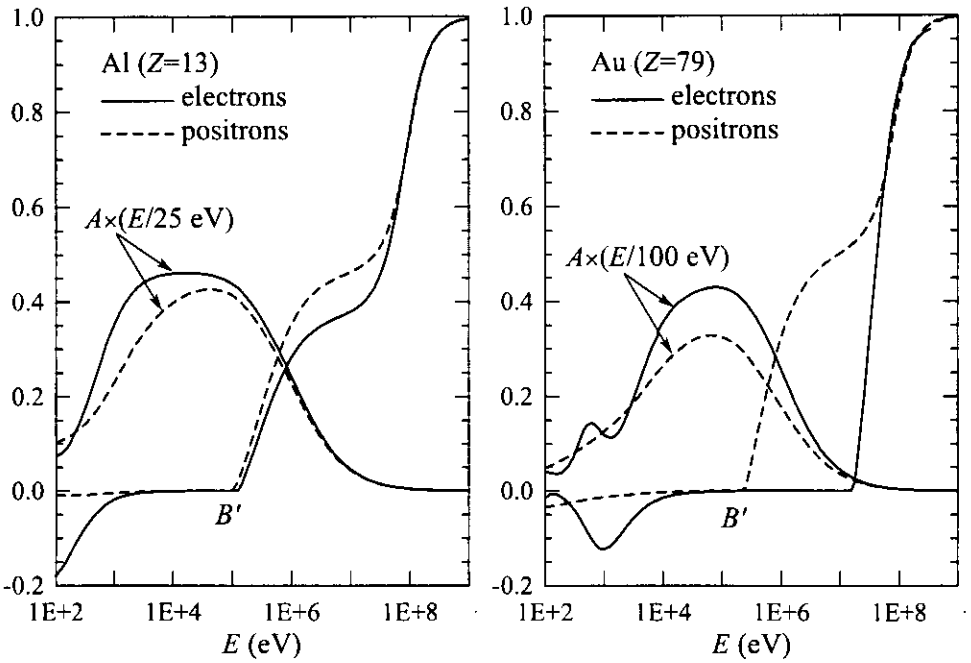


Figure 3.5: Parameters of the MW model for scattering of electrons and positrons by aluminium and gold atoms. The scale of the energy axes is logarithmic.

partial-wave DCS is much narrower than the Wentzel distribution.

The MW model is directly applicable to compounds (and mixtures) by using the appropriate values of the total cross section and the first and second transport cross sections. In PENELOPE, these are calculated from atomic total and transport cross sections by means of the additivity approximation (incoherent sum of scattered intensities). This amounts to neglecting chemical binding effects. A more accurate approach, which yields a good estimate of these effects, is provided by the following independent atom approximation (Walker, 1968; Yates, 1968). Assume that the interaction of the projectile with each atom is still given by the free-atom static potential (3.4). The molecular DCS may then be evaluated by adding the waves (not the currents) scattered from the various atoms in the molecule and averaging over molecular orientations. The resulting DCS is given by

$$\frac{\mathrm{d}\sigma_{\mathrm{el}}}{\mathrm{d}\Omega} = \sum_{i,j} \frac{\sin(qa_{ij}/\hbar)}{qa_{ij}/\hbar} \left[f_i(\theta) f_j^*(\theta) + g_i(\theta) g_j^*(\theta) \right], \tag{3.30}$$

where $q = 2\hbar k \sin(\theta/2)$ is the momentum transfer, a_{ij} is the distance between the atoms i and j and f_i, g_i are the scattering amplitudes, eq. (3.6), for the atom i. It has been claimed that DCSs obtained from this formulation agree with experiments to within ~2% (Walker, 1968; Yates, 1968). DCSs for scattering of 100 eV and 2.5 keV electrons in water vapour, obtained from the simple additivity rule and computed from

eq. (3.30), are compared in fig. 3.6. It is seen that, for energies above a few keV, chemical binding causes a slight distortion of the DCS at small angles, and a slight rippling for intermediate angles. Therefore, the use of the additivity approximation (i.e. neglecting chemical binding effects) in Monte Carlo simulation at these energies is justified.

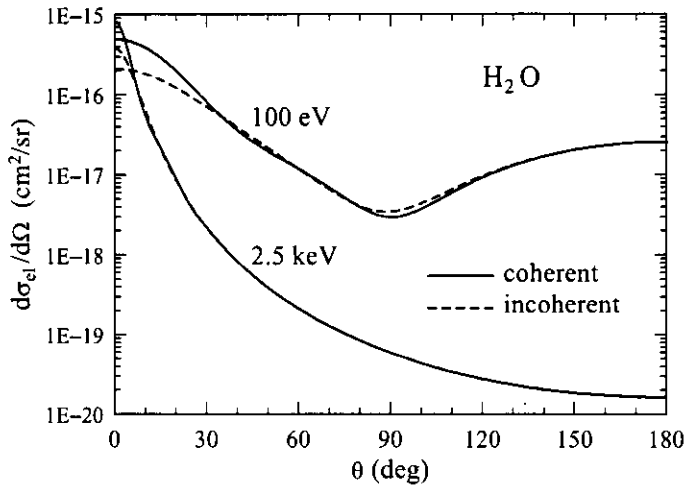


Figure 3.6: Differential cross sections for elastic scattering of electrons by water molecules, calculated as the coherent sum of scattered waves, eq. (3.30), and from the additivity approximation (incoherent sum).

3.1.2 Simulation of single elastic events with the MW model

As mentioned above, the angular distribution in single elastic events is axially symmetrical about the direction of incidence. Hence, the azimuthal scattering angle ϕ is sampled uniformly in the interval $(0, 2\pi)$ using the sampling formula $\phi = 2\pi\xi$. In detailed simulations, μ is sampled in the whole interval (0,1). However, we shall also make use of the MW model for mixed simulation (see chapter 4), in which only hard events, with deflection μ larger than a given cutoff value μ_c, are sampled individually. In this section we describe analytical (i.e. exact) methods for random sampling of μ in the restricted interval $(\mu_c, 1)$.

- Case I. The cumulative distribution function of $p_{\mathrm{MW,I}}(\mu)$ is

$$\mathcal{P}_{\mathrm{MW,I}}(\mu) \equiv \int_0^{\mu} p_{\mathrm{MW,I}}(\mu')\,\mathrm{d}\mu' = \begin{cases} (1-B)\dfrac{(1+A)\mu}{A+\mu} & \text{if } 0 \le \mu < \langle\mu\rangle, \\ B + (1-B)\dfrac{(1+A)\mu}{A+\mu} & \text{if } \langle\mu\rangle \le \mu \le 1. \end{cases} \tag{3.31}$$

Owing to the analytical simplicity of this function, the random sampling of μ can be performed by using the inverse transform method (section 1.2.2). The sampling equation

for μ in (0,1) reads

$$\mu = \mathcal{P}^{-1}_{\mathrm{MW,I}}(\xi), \tag{3.32}$$

where $\mathcal{P}^{-1}_{\mathrm{MW,I}}(\xi)$ is the inverse of the cumulative distribution function, which is given by

$$\mathcal{P}^{-1}_{\mathrm{MW,I}}(\xi) = \begin{cases} \dfrac{\xi A}{(1-B)(1+A)-\xi} & \text{if } 0 \le \xi < \xi_0, \\ \langle\mu\rangle & \text{if } \xi_0 \le \xi < \xi_0 + B, \\ \dfrac{(\xi - B)A}{(1-B)(1+A)-(\xi-B)} & \text{if } \xi_0 + B \le \xi \le 1, \end{cases} \tag{3.33}$$

with

$$\xi_0 = (1-B)\frac{(1+A)\langle\mu\rangle}{A+\langle\mu\rangle}. \tag{3.34}$$

To sample μ in the restricted interval $(\mu_c,1)$, we can still use the inverse transform method, ec. (3.32), but with the random number ξ sampled uniformly in the interval $(\xi_c,1)$ with

$$\xi_c = \mathcal{P}_{\mathrm{MW,I}}(\mu_c). \tag{3.35}$$

- Case II. The cumulative distribution function is

$$\begin{aligned} \mathcal{P}_{\mathrm{MW,II}}(\mu) &\equiv \int_0^\mu p_{\mathrm{MW,II}}(\mu')\,\mathrm{d}\mu' \\ &= \begin{cases} (1-B)\dfrac{(1+A)\mu}{A+\mu} & \text{if } 0 \le \mu < \frac{1}{2}, \\ (1-B)\dfrac{(1+A)\mu}{A+\mu} + B\,4\left[\mu^2 - \mu + \dfrac{1}{4}\right] & \text{if } \frac{1}{2} \le \mu \le 1. \end{cases} \end{aligned} \tag{3.36}$$

In principle, to sample μ in (0,1), we can adopt the inverse transform method. The sampling equation

$$\xi = \mathcal{P}_{\mathrm{MW,II}}(\mu) \tag{3.37}$$

can be cast in the form of a cubic equation. This equation can be solved either by using the analytical solution formulas for the cubic equation, which are somewhat complicated, or numerically, e.g. by the Newton-Raphson method. We employ this last procedure to determine the cutoff deflection (see section 4.1) for mixed simulation. To sample μ in the restricted interval $(\mu_c,1)$ we use the composition method, which is easier than solving eq. (3.37). Notice that the sampling from the (restricted) Wentzel and from the triangle distributions can be performed analytically by the inverse transform method.

3.2 Inelastic collisions

The dominant energy loss mechanisms for electrons and positrons with intermediate and low energies are inelastic collisions, i.e. interactions that produce electronic excitations

and ionizations in the medium. The quantum theory of inelastic collisions of charged particles with individual atoms and molecules was first formulated by Bethe (1930, 1932) on the basis of the first-order (plane-wave) Born approximation. The extension of the theory to inelastic collisions in condensed materials has been discussed by Fano (1963). The formal aspects of the quantum theory for condensed matter are quite complicated. Fortunately, the results are essentially equivalent to those from classical dielectric theory.

The effect of individual inelastic collisions *on the projectile* is completely specified by giving the energy loss W and the polar and azimuthal scattering angles θ and ϕ, respectively. For amorphous media with randomly oriented atoms (or molecules), the DCS for inelastic collisions is independent of the azimuthal scattering angle ϕ. Instead of the polar scattering angle θ, it is convenient to use the recoil energy Q [see eqs. (A.29) and (A.30)], defined by

$$Q(Q + 2\mathrm{m_e}c^2) = (cq)^2. \tag{3.38}$$

The quantity q is the magnitude of the momentum transfer $\mathbf{q} \equiv \mathbf{p} - \mathbf{p}'$, where $\mathbf{p}$ and $\mathbf{p}'$ are the linear momenta of the projectile before and after the collision. Notice that Q is the kinetic energy of an electron that moves with a linear momentum equal to q.

Let us first consider the inelastic interactions of electrons or positrons ($z_0^2 = 1$) with an isolated atom (or molecule) containing Z electrons in its ground state. The DCS for collisions with energy loss W and recoil energy Q, obtained from the first Born approximation, can be written in the form (Fano, 1963)

$$\frac{\mathrm{d}^2\sigma_{\mathrm{in}}}{\mathrm{d}W\,\mathrm{d}Q} = \frac{2\pi z_0^2 e^4}{\mathrm{m_e}v^2}\left(\frac{2\mathrm{m_e}c^2}{WQ(Q+2\mathrm{m_e}c^2)} + \frac{\beta^2\sin^2\theta_{\mathrm{r}} W 2\mathrm{m_e}c^2}{[Q(Q+2\mathrm{m_e}c^2) - W^2]^2}\right)\frac{\mathrm{d}f(Q,W)}{\mathrm{d}W}, \tag{3.39}$$

where $v = \beta c$ is the velocity of the projectile. θ_{r} is the angle between the initial momentum of the projectile and the momentum transfer, which is given by eq. (A.42),

$$\cos^2\theta_{\mathrm{r}} = \frac{W^2/\beta^2}{Q(Q+2\mathrm{m_e}c^2)}\left(1 + \frac{Q(Q+2\mathrm{m_e}c^2) - W^2}{2W(E+\mathrm{m_e}c^2)}\right)^2. \tag{3.40}$$

The result (3.39) is obtained in the Coulomb gauge (Fano, 1963); the two terms on the right-hand side are the contributions from interactions through the instantaneous (longitudinal) Coulomb field and through the exchange of virtual photons (transverse field), respectively. The factor $\mathrm{d}f(Q,W)/\mathrm{d}W$ is the atomic generalized oscillator strength (GOS), which completely determines the effect of inelastic interactions *on the projectile*, within the Born approximation. Notice, however, that knowledge of the GOS does not suffice to describe the energy spectrum and angular distribution of secondary knock-on electrons (delta rays).

The GOS can be represented as a surface over the (Q,W) plane, which is called the Bethe surface (see Inokuti, 1971; Inokuti et al., 1978). Unfortunately, the GOS is known in analytical form only for two simple systems, namely, the (non-relativistic) hydrogenic ions (see fig. 3.7) and the free-electron gas. Even in these cases, the analytical expressions of the GOSs are too complicated for simulation purposes. For ionization of inner shells, the GOS can be computed numerically from first principles (see e.g. Manson, 1972), but

using GOSs defined through extensive numerical tables is impractical for Monte Carlo simulation. Fortunately, the physics of inelastic collisions is largely determined by a few global features of the Bethe surface. Relatively simple GOS models can be devised that are consistent with these features and, therefore, lead to a fairly realistic description of inelastic interactions (see e.g. Salvat and Fernández-Varea, 1992).

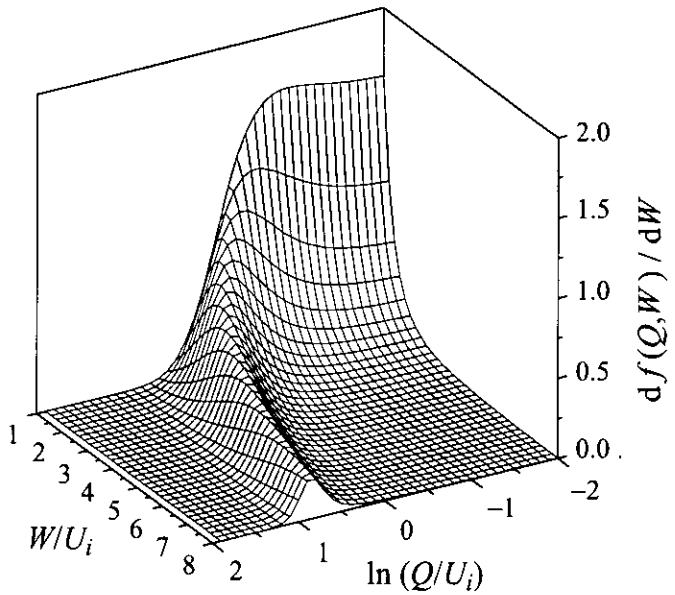


Figure 3.7: The GOS for ionization of the hydrogen atom ($Z = 1$) in the ground state. All energies are in units of the ionization energy $U_i = 13.6$ eV. The GOS for ionization of (non-relativistic) hydrogenic ions is independent of Z if energies are expressed in units of the ionization energy.

As mentioned above, the "atomic" DCS for inelastic interactions in dense media can be obtained from a semiclassical treatment in which the medium is considered as a dielectric, characterized by a complex dielectric function $\epsilon(k,\omega)$, which depends on the wave number k and the frequency ω. In the classical picture, the (external) electric field of the projectile polarizes the medium producing an induced electric field that causes the slowing down of the projectile. The dielectric function relates the Fourier components of the total (external+induced) and the external electric potentials. It is convenient to interpret the quantities $q = \hbar k$ and $W = \hbar\omega$ as the momentum and energy transfers and consider that the dielectric function depends on the variables Q [defined by eq. (3.38)] and W. The DCSs obtained from the dielectric and quantum treatments are consistent (i.e. the former reduces to the latter for a low-density medium) if one assumes the identity

$$\frac{\mathrm{d}f(Q,W)}{\mathrm{d}W} \equiv W\frac{Q+\mathrm{m_e}c^2}{\mathrm{m_e}c^2}\frac{2Z}{\pi\Omega_\mathrm{p}^2}\,\mathrm{Im}\left(\frac{-1}{\epsilon(Q,W)}\right), \tag{3.41}$$

 3518

where Ω_p is the plasma energy of a free-electron gas with the electron density of the medium, given by

$$\Omega_p^2 = 4\pi \mathcal{N} Z \hbar^2 e^2 / m_e. \tag{3.42}$$

Eq. (3.41) establishes the connection between the atomic GOS (a property of individual atoms) and the dielectric function (a macroscopic concept). The DCS for the condensed medium can be expressed in the form [cf. eq. (3.39)],

$$\frac{d^2\sigma_{in}}{dW\,dQ} = \frac{2\pi z_0^2 e^4}{m_e v^2} \frac{df(Q,W)}{dW} \left(\frac{2m_e c^2}{WQ(Q+2m_e c^2)} + \left\{ \frac{\beta^2 \sin^2\theta_r W 2m_e c^2}{[Q(Q+2m_e c^2) - W^2]^2} - \mathcal{D}(Q,W) \right\} \right), \tag{3.43}$$

where the term $\mathcal{D}(Q,W)$, which is appreciable only for small Q, accounts for the so-called density-effect correction (Sternheimer, 1952). The origin of this term is the polarizability of the medium, which "screens" the distant transverse interaction causing a net reduction of its contribution to the stopping power. The density-effect correction $\mathcal{D}(Q,W)$ is determined by the dielectric function that, in turn, is related to the GOS. Thus, the GOS contains all the information needed to compute the DCS for electron/positron inelastic interactions in condensed media.

In the limit of very large recoil energies, the binding and momentum distribution of the target electrons have a small effect on the interaction. Therefore, in the large-Q region, the target electrons behave as if they were essentially free and at rest and, consequently, the GOS reduces to a ridge along the line $W = Q$, which was named the Bethe ridge by Inokuti (1971). In the case of hydrogenic ions in the ground state, fig. 3.7, the Bethe ridge becomes clearly visible at relatively small recoil energies, of the order of the ionization energy U_i. For smaller Q's, the structure of the Bethe surface is characteristic of the material. In the limit $Q \to 0$, the GOS reduces to the optical oscillator strength (OOS),

$$\frac{df(W)}{dW} \equiv \frac{df(Q=0,W)}{dW}, \tag{3.44}$$

which is closely related to the (dipole) photoelectric cross section for photons of energy W (Fano, 1963). Experimental information on the OOS is provided by measurements of either photoelectric cross sections or dielectric functions (see e.g. Fernández-Varea et al., 1993a). The GOS satisfies the Bethe sum rule (Inokuti, 1971)

$$\int_0^\infty \frac{df(Q,W)}{dW}\,dW = Z \qquad \text{for any } Q. \tag{3.45}$$

This sum rule, which is a result from non-relativistic theory (see e.g. Mott and Massey, 1965), is assumed to be generally satisfied. It leads to the interpretation of the GOS as the effective number of electrons per unit energy transfer that participate in interactions with given recoil energy Q. The mean excitation energy I, defined by (Fano, 1963; Inokuti, 1971)

$$Z \ln I = \int_0^\infty \ln W \frac{df(W)}{dW}\,dW, \tag{3.46}$$

 8519

plays a central role in the Bethe stopping power formula [eq. (3.105) below]. This quantity has been determined empirically for a large number of materials (see Berger and Seltzer, 1982, and references therein) from measurements of the stopping power of heavy charged particles and/or from experimental optical dielectric functions. In the following, we shall assume that the mean excitation energy of the stopping medium is known.

3.2.1 GOS model

The simulation of inelastic collisions of electrons and positrons in PENELOPE is performed on the basis of the following GOS model, which is tailored to allow fast random sampling of W and Q. We assume that the GOS splits into contributions from the different atomic electron shells. Each atomic shell k is characterized by the number f_k of electrons in the shell and the ionization energy U_k. To model the contribution of a shell to the GOS, we refer to the example of the hydrogen atom (fig. 3.7) and observe that for $Q > U_k$ the GOS reduces to the Bethe ridge, whereas for $Q < U_k$ it is nearly constant with Q and decreases rapidly with W; a large fraction of the OOS concentrates in a relatively narrow W-interval. Consideration of other well-known systems, such as inner shells of heavy atoms (Manson, 1972) and the free-electron gas (Lindhard and Winther, 1964), shows that these gross features of the GOS are universal. Liljequist (1983) proposed modelling the GOS of each shell by means of a "δ-oscillator", which is an entity with a simple excitation spectrum given by (see fig. 3.8)

$$\frac{\mathrm{d}f_k(Q,W)}{\mathrm{d}W} = f_k\left[\delta(W-W_k)\Theta(W_k-Q)+\delta(W-Q)\Theta(Q-W_k)\right]. \qquad (3.47)$$

The first term represents resonant low-Q (distant) interactions, which are described as a single resonance at the energy W_k. The second term corresponds to large-Q (close) interactions, in which the target electrons react as if they were free and at rest ($W = Q$). Notice that the oscillator GOS satisfies the sum rule

$$\int_0^\infty \frac{\mathrm{d}f_k(Q,W)}{\mathrm{d}W}\,\mathrm{d}W = f_k \qquad \text{for any } Q \qquad (3.48)$$

and, consequently, the oscillator strength f_k can be identified with the number of electrons represented by the oscillator. Here we shall adopt Liljequist's model and describe the excitations of a shell (or a group of shells) by means of a single oscillator, with the resonance energy considered as an adjustable parameter. The oscillator model gives a Bethe ridge with zero width, i.e. the broadening caused by the momentum distribution of the target electrons is neglected. This is not a serious drawback for light projectiles (electrons and positrons), but it can introduce sizeable errors in the computed cross sections for slow heavy projectiles with $m \gg \mathrm{m_e}$. Our oscillator model also disregards the fact that, for low-Q interactions, there is a transfer of oscillator strength from inner to outer shells (see e.g. Shiles et al., 1980).

In mixed (class II) simulations, only hard collisions, with energy loss larger than a specified cutoff value W_{cc}, are simulated (see chapter 4). The effect of soft interactions

 3620

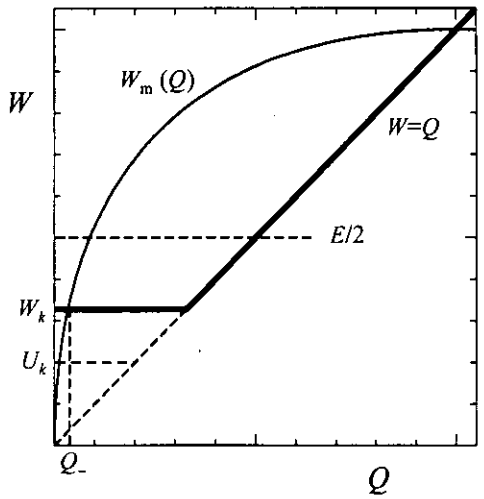


Figure 3.8: Oscillator model for the GOS of an inner shell with $U_k = 2$ keV. The continuous curve represents the maximum allowed energy loss as a function of the recoil energy, $W_m(Q)$, for electrons/positrons with $E = 10$ keV. For distant interactions the possible recoil energies lie in the interval from Q_- to W_k. Recoil energies larger than W_k correspond to close interactions. The largest allowed energy loss W_{max} is $E/2$ for electrons and E for positrons (see text).

(with $W < W_{cc}$) is described by means of a multiple scattering approximation, which does not require detailed knowledge of the shell DCSs. K and L shells with ionization energy U_i larger than max(200 eV, W_{cc}) will be referred to as "inner" shells. The partial GOS of an inner shell will be represented by the oscillator

$$\frac{\mathrm{d}f_i(Q,W)}{\mathrm{d}W} = f_i\left[\delta(W-W_i)\Theta(W_i-Q) + \delta(W-Q)\Theta(Q-W_i)\right], \qquad (3.49)$$

with resonance energy $W_i = 1.65\ U_i$. This value of W_i is derived from the hydrogenic model (see below), which is expected to be approximately valid for the innermost shells. It should be noted that the shell ionization cross sections obtained from this model are only roughly approximate. Their use in a Monte Carlo code is permissible only because ionization of inner shells is a low-probability process (see fig. 3.9 below) that has a weak effect on the global transport properties. In cases where inner-shell ionization is directly observed (e.g. in the simulation of x-ray emission by electron bombardment), a more accurate description of the process should be used.

The largest contribution to the total cross section arises from low-W (soft) excitations. Therefore, the total cross section is mostly determined by the OOS of weakly bound electrons, which is strongly dependent on the state of aggregation. In the case of conductors and semiconductors, electrons in the outermost shells form the conduction band (cb). These electrons can move quite freely through the medium and, hence, their binding energy is set to zero, $U_{cb} = 0$. Excitations of the conduction band will be

described by a single oscillator,

$$\frac{\mathrm{d}f_{\mathrm{cb}}(Q,W)}{\mathrm{d}W} = f_{\mathrm{cb}}\left[\delta(W-W_{\mathrm{cb}})\Theta(W_{\mathrm{cb}}-Q)+\delta(W-Q)\Theta(Q-W_{\mathrm{cb}})\right]. \tag{3.50}$$

The parameters W_{cb} and f_{cb} should be identified with the plasmon energy and the effective number of electrons (per atom or molecule) that participate in plasmon excitations; these quantities can be deduced e.g. from electron energy-loss spectra or from measured optical data. When this information is not available, we will simply fix the value of f_{cb} (as the number of electrons with ionization energies less than, say, 15 eV) and set the resonance energy W_{cb} equal to the plasmon energy of a free-electron gas with the same density as that of conduction electrons,

$$W_{\mathrm{cb}} = \sqrt{4\pi\mathcal{N}f_{\mathrm{cb}}\hbar^2e^2/\mathrm{m_e}} = \sqrt{\frac{f_{\mathrm{cb}}}{Z}}\,\Omega_{\mathrm{p}}. \tag{3.51}$$

This is a fairly realistic model for free-electron-like metals (such as aluminium), because the resonance energy is set equal to the plasmon energy of a free-electron gas (see e.g. Kittel, 1976).

Electron shells other than K and L shells, or with $U_j < \max(200\ \mathrm{eV}, W_{\mathrm{cc}})$, will be referred to as "outer" shells. In the case of conductors, shells that contribute to the conduction band are excluded from the set of outer shells. Each outer shell is described by an oscillator

$$\frac{\mathrm{d}f_j(Q,W)}{\mathrm{d}W} = f_j\left[\delta(W-W_j)\Theta(W_j-Q)+\delta(W-Q)\Theta(Q-W_j)\right], \tag{3.52}$$

with resonance energy $W_j = gU_j$, where g is an empirical adjustment factor (the same for all outer shells) which will be determined below. We thus arrive at the following expression for the GOS,

$$\begin{aligned}\frac{\mathrm{d}f(Q,W)}{\mathrm{d}W} &= \sum_k \frac{\mathrm{d}f_k(Q,W)}{\mathrm{d}W} \\ &= \sum_k f_k\left[\delta(W-W_k)\Theta(W_k-Q)+\delta(W-Q)\Theta(Q-W_k)\right],\end{aligned} \tag{3.53}$$

where the summation in k extends over all inner and outer electron shells (and the conduction band, in the case of conductors). This GOS model satisfies the Bethe sum rule (3.45),

$$\int_0^\infty \frac{\mathrm{d}f(Q,W)}{\mathrm{d}W}\,\mathrm{d}W = \sum_k f_k = Z \qquad \text{for any } Q. \tag{3.54}$$

The corresponding OOS reduces to

$$\frac{\mathrm{d}f(Q=0,W)}{\mathrm{d}W} = \sum_k f_k\delta(W-W_k), \tag{3.55}$$

 3522

which has the same form (a superposition of resonances) as the OOS used by Sternheimer (1952) in his calculations of the density effect correction. In order to reproduce the high-energy stopping power given by the Bethe formula (Berger and Seltzer, 1982), the excitation energies and oscillator strengths must lead, through eq. (3.46), to the accepted value of the mean excitation energy I, i.e.

$$\sum_k f_k \ln W_k = Z \ln I. \tag{3.56}$$

Following Sternheimer (1952), we use this relation to determine the empirical adjustment factor g $(= W_j/U_j)$,

$$\ln g = \left[Z \ln I - \sum_i f_i \ln W_i - f_{\rm cb} \ln W_{\rm cb} - \sum_j f_j \ln U_j \right] \left(\sum_j f_j \right)^{-1}. \tag{3.57}$$

For a one-shell system, such as the hydrogen atom, the constraint (3.56) implies that the resonance energy W_i is equal to I [see the definition (3.46)]. Considering the $\sim W^{-3}$ dependence of the hydrogenic OOS, it is concluded that g should be of the order of $\exp(1/2) = 1.65$ (Sternheimer et al., 1982). It is worth noting that the Sternheimer adjustment factor g is a property of the considered medium; therefore, the DCSs for ionization of an outer shell of a given element in two different compounds may be slightly different.

The present theory is directly applicable to compounds (and mixtures), since the oscillators may pertain either to atoms or molecules. When the value of the mean excitation energy of the compound is not known, it may be estimated from Bragg's additivity rule as follows. Consider a compound $\mathrm{X_xY_y}$, in which the molecules consist of x atoms of the element X and y atoms of the element Y. The number of electrons per molecule is $Z_{\rm M} = \mathrm{x}Z_{\rm X} + \mathrm{y}Z_{\rm Y}$, where $Z_{\rm X}$ stands for the atomic number of element X. According to the additivity rule, the GOS of the compound is approximated as the sum of the atomic GOSs of the atoms so that

$$Z_{\rm M} \ln I = \mathrm{x}Z_{\rm X} \ln I_{\rm X} + \mathrm{y}Z_{\rm Y} \ln I_{\rm Y}, \tag{3.58}$$

where $I_{\rm X}$ denotes the mean excitation energy of element X.

For heavy elements, and also for compounds and mixtures with several elements, the number of electron shells may be fairly large (of the order of sixty for an alloy of two heavy metals). In these cases, it would be impractical to treat all shells with the same detail/accuracy. In fact, the description of the outer shells can be simplified without sacrificing the reliability of the simulation results. In PENELOPE, the maximum number of electron shells for each material is limited. When the number of actual shells is too large, outer shells that have similar resonance energies are grouped together and described by a single oscillator. The grouping is made in such a way that the contribution of the group oscillator to the mean excitation energy I equals the sum of contributions of the grouped oscillators; this ensures that grouping will not alter the stopping power of fast particles (with E substantially greater than the ionization energy of the grouped oscillators).

3.2.2 Differential cross sections

The DCS for inelastic collisions obtained from our GOS model can be split into contributions from distant logitudinal, distant transverse and close interactions,

$$\frac{\mathrm{d}^2\sigma_{\mathrm{in}}}{\mathrm{d}W\,\mathrm{d}Q} = \frac{\mathrm{d}^2\sigma_{\mathrm{dis,l}}}{\mathrm{d}W\,\mathrm{d}Q} + \frac{\mathrm{d}^2\sigma_{\mathrm{dis,t}}}{\mathrm{d}W\,\mathrm{d}Q} + \frac{\mathrm{d}^2\sigma_{\mathrm{clo}}}{\mathrm{d}W\,\mathrm{d}Q}. \tag{3.59}$$

The DCS for distant longitudial interactions is given by the first term in eq. (3.43),

$$\frac{\mathrm{d}^2\sigma_{\mathrm{dis,l}}}{\mathrm{d}W\,\mathrm{d}Q} = \frac{2\pi e^4}{\mathrm{m_e}v^2}\sum_k f_k \frac{1}{W}\frac{2\mathrm{m_e}c^2}{Q(Q+2\mathrm{m_e}c^2)}\delta(W-W_k)\,\Theta(W_k-Q). \tag{3.60}$$

As mentioned above, the DCS for distant transverse interactions has a complicated expression. To simplify it, we shall ignore the (very small) angular deflections of the projectile in these interactions and replace the expression in curly brackets in eq. (3.43) by an averaged W-independent value that gives the exact contribution of the distant transverse interactions to the high-energy stopping power (Salvat and Fernández-Varea, 1992). This yields the following approximate expression for the DCS of distant transverse interactions,

$$\begin{aligned}\frac{\mathrm{d}^2\sigma_{\mathrm{dis,t}}}{\mathrm{d}W\,\mathrm{d}Q} &= \frac{2\pi e^4}{\mathrm{m_e}v^2}\sum_k f_k \frac{1}{W}\left\{\ln\left(\frac{1}{1-\beta^2}\right) - \beta^2 - \delta_{\mathrm{F}}\right\} \\ &\quad\times \delta(W-W_k)\,\Theta(W_k-Q_-)\,\delta(Q-Q_-),\end{aligned} \tag{3.61}$$

where Q_- is the minimum recoil energy[3] for the energy transfer W, eq. (A.31), and δ_{F} is the Fermi density effect correction on the stopping power, which has been studied extensively in the past (Sternheimer, 1952; Fano, 1963). δ_{F} can be computed as (Fano, 1963)

$$\delta_{\mathrm{F}} \equiv \frac{1}{Z}\int_0^\infty \frac{\mathrm{d}f(Q=0,W)}{\mathrm{d}W}\ln\left(1+\frac{L^2}{W^2}\right)\mathrm{d}W - \frac{L^2}{\Omega_{\mathrm{p}}^2}\left(1-\beta^2\right), \tag{3.62}$$

where L is a real-valued function of β^2 defined as the positive root of the following equation (Inokuti and Smith, 1982):

$$\mathcal{F}(L) \equiv \frac{1}{Z}\Omega_{\mathrm{p}}^2\int_0^\infty \frac{1}{W^2+L^2}\frac{\mathrm{d}f(Q=0,W)}{\mathrm{d}W}\mathrm{d}W = 1-\beta^2. \tag{3.63}$$

The function $\mathcal{F}(L)$ decreases monotonically with L, and hence, the root $L(\beta^2)$ exists only when $1-\beta^2 < \mathcal{F}(0)$; otherwise it is $\delta_{\mathrm{F}} = 0$. Therefore, the function $L(\beta^2)$ starts with zero at $\beta^2 = 1-\mathcal{F}(0)$ and grows monotonically with increasing β^2. With the OOS, given by eq. (3.53), we have

$$\mathcal{F}(L) = \frac{1}{Z}\Omega_{\mathrm{p}}^2\sum_k \frac{f_k}{W_k^2+L^2} \tag{3.64}$$

[3]The recoil energy Q_- corresponds to $\theta = 0$, i.e. we consider that the projectile is not deflected by distant transverse interactions.

and

$$\delta_{\mathrm{F}} \equiv \frac{1}{Z}\sum_k f_k \ln\left(1+\frac{L^2}{W_k^2}\right) - \frac{L^2}{\Omega_{\mathrm{p}}^2}\left(1-\beta^2\right). \tag{3.65}$$

In the high-energy limit ($\beta \to 1$), the L value resulting from eq. (3.63) is large ($L \gg W_k$) and can be approximated as $L^2 = \Omega_{\mathrm{p}}^2/(1-\beta^2)$. Then, using the Bethe sum rule ($\sum f_k = Z$) and the relation (3.56), we obtain

$$\delta_{\mathrm{F}} \simeq \ln\left(\frac{\Omega_{\mathrm{p}}^2}{(1-\beta^2)I^2}\right) - 1, \qquad \text{when } \beta \to 1. \tag{3.66}$$

The DCS for close collisions is given by

$$\frac{\mathrm{d}^2\sigma_{\mathrm{clo}}}{\mathrm{d}W\,\mathrm{d}Q} = \frac{2\pi e^4}{\mathrm{m_e}v^2}\sum_k f_k \frac{1}{W}\left(\frac{2\mathrm{m_e}c^2}{W(W+2\mathrm{m_e}c^2)} + \frac{\beta^2\sin^2\theta_{\mathrm{clo}}}{2\mathrm{m_e}c^2}\right)\delta(W-Q)\,\Theta(W-W_k),$$

where θ_{clo} is the recoil angle, defined by eq. (3.40) with $Q = W$,

$$\cos^2\theta_{\mathrm{clo}} = \frac{W}{E}\,\frac{E+2\mathrm{m_e}c^2}{W+2\mathrm{m_e}c^2}. \tag{3.67}$$

We have

$$\frac{\mathrm{d}^2\sigma_{\mathrm{clo}}}{\mathrm{d}W\,\mathrm{d}Q} = \frac{2\pi e^4}{\mathrm{m_e}v^2}\sum_k f_k \frac{1}{W^2}\left(1+\frac{\beta^2(E-W)W - EW}{E(W+2\mathrm{m_e}c^2)}\right)\delta(W-Q)\,\Theta(W-W_k). \tag{3.68}$$

DCS for close collisions of electrons

When the projectile is an electron, the DCS must be corrected to account for the indistinguishability of the projectile and the target electrons. For distant interactions, the effect of this correction is small (much smaller than the distortion introduced by our modelling of the GOS) and will be neglected. The energy loss DCS for binary collisions of electrons with free electrons at rest, obtained from the Born approximation with proper account of exchange, is given by the Møller (1932) formula,

$$\begin{aligned}\frac{\mathrm{d}^2\sigma_{\mathrm{M}}}{\mathrm{d}W\mathrm{d}Q} = \frac{2\pi e^4}{\mathrm{m_e}v^2}\frac{1}{W^2}\Bigg[1 + \left(\frac{W}{E-W}\right)^2 - \frac{W}{E-W} \\ + a\left(\frac{W}{E-W} + \frac{W^2}{E^2}\right)\Bigg]\delta(W-Q),\end{aligned} \tag{3.69}$$

where

$$a = \left(\frac{E}{E+\mathrm{m_e}c^2}\right)^2 = \left(\frac{\gamma-1}{\gamma}\right)^2. \tag{3.70}$$

To introduce exchange effects in the DCS for close interactions of electrons, we replace the factor in parentesis in eq. (3.68) by the analogous factor in Møller's formula , i.e. we take

$$\frac{\mathrm{d}^2\sigma_{\mathrm{clo}}^{(-)}}{\mathrm{d}W\,\mathrm{d}Q} = \frac{2\pi e^4}{\mathrm{m_e}v^2}\sum_k f_k \frac{1}{W^2}F^{(-)}(E,W)\delta(W-Q)\,\Theta(W-W_k), \tag{3.71}$$

with

$$F^{(-)}(E,W) \equiv 1 + \left(\frac{W}{E-W}\right)^2 - \frac{W}{E-W} + a\left(\frac{W}{E-W} + \frac{W^2}{E^2}\right). \quad (3.72)$$

In the final state we have two indistinguishable free electrons, and it is natural to consider the fastest one as the "primary". Accordingly, the maximum allowed energy transfer in close collisions is

$$W_{\max} = E/2. \quad (3.73)$$

After ionizing an inner shell i, the primary electron has a kinetic energy $E - W$, the "secondary" electron (delta ray) is ejected with kinetic energy $W - U_i$, and the residual ion is left in an excited state, with a vacancy in the inner shell i, which corresponds to an excitation energy equal to U_i. This energy is released by emission of x rays and Auger electrons (see section 2.6). When the ionization occurs in an outer shell or in the conduction band, the initial energy of the secondary electron is set equal to W and no fluorescent radiation from the ionized atom is followed by the simulation program. This is equivalent to assuming that the secondary electron carries away the excitation energy of the target atom.

DCS for close collisions of positrons

Positrons in matter are unstable particles that annihilate with electrons giving photons (see section 3.4). On the other hand, electron-positron pairs can be created if enough electromagnetic energy ($> 2m_e c^2$) is available (either from real or virtual photons). A positron does not interact with matter as a usual (stable) positively charged particle, since the competing process of annihilation followed by re-creation can cause the same transitions as "direct" scattering (see e.g. Sakurai, 1967). The DCS for binary collisions of positrons with free electrons at rest, obtained from the first Born approximation including the "annihilation/creation" mechanism, is given by the Bhabha (1936) formula,

$$\frac{d^2\sigma_B}{dW dQ} = \frac{2\pi e^4}{m_e v^2}\frac{1}{W^2}\left[1 - b_1\frac{W}{E} + b_2\left(\frac{W}{E}\right)^2 - b_3\left(\frac{W}{E}\right)^3 + b_4\left(\frac{W}{E}\right)^4\right]\delta(W-Q), \quad (3.74)$$

where

$$b_1 = \left(\frac{\gamma-1}{\gamma}\right)^2 \frac{2(\gamma+1)^2-1}{\gamma^2-1}, \qquad b_2 = \left(\frac{\gamma-1}{\gamma}\right)^2 \frac{3(\gamma+1)^2+1}{(\gamma+1)^2},$$

$$b_3 = \left(\frac{\gamma-1}{\gamma}\right)^2 \frac{2\gamma(\gamma-1)}{(\gamma+1)^2}, \qquad b_4 = \left(\frac{\gamma-1}{\gamma}\right)^2 \frac{(\gamma-1)^2}{(\gamma+1)^2}. \quad (3.75)$$

To account approximately for the effect of annihilation/creation on the DCS for close inelastic interactions of positrons, we shall use the expression (3.68), with the factor in parenthesis replaced by the Bhabha factor,

$$F^{(+)}(E,W) = 1 - b_1\frac{W}{E} + b_2\left(\frac{W}{E}\right)^2 - b_3\left(\frac{W}{E}\right)^3 + b_4\left(\frac{W}{E}\right)^4. \quad (3.76)$$

That is,

$$\frac{\mathrm{d}^2\sigma_{\mathrm{clo}}^{(+)}}{\mathrm{d}W\,\mathrm{d}Q} = \frac{2\pi e^4}{\mathrm{m_e}v^2}\sum_k f_k \frac{1}{W^2} F^{(+)}(E,W)\delta(W-Q)\,\Theta(W-W_k). \tag{3.77}$$

The ionization process is simulated as in the case of electrons. In the ionization of an inner shell i, the delta ray is emitted with initial kinetic energy equal to $W-U_i$ and the subsequent relaxation of the residual ion is followed. When the ionization is in an outer shell (or in the conduction band), the secondary electron is assumed to be ejected with kinetic energy W and no fluorescent radiation is followed. Notice that the maximum energy loss in collisions of positrons with energy E is $W_{\mathbf{max}} = E$.

3.2.3 Integrated cross sections

The energy-loss DCS is defined as

$$\frac{\mathrm{d}\sigma_{\mathrm{in}}}{\mathrm{d}W} \equiv \int_{Q_-}^{Q_+} \frac{\mathrm{d}^2\sigma_{\mathrm{in}}}{\mathrm{d}W\,\mathrm{d}Q}\,\mathrm{d}Q = \frac{\mathrm{d}\sigma_{\mathrm{dis,l}}}{\mathrm{d}W} + \frac{\mathrm{d}\sigma_{\mathrm{dis,t}}}{\mathrm{d}W} + \frac{\mathrm{d}\sigma_{\mathrm{clo}}}{\mathrm{d}W}, \tag{3.78}$$

where Q_- and Q_+ are the minimum and maximum kinematically allowed recoil energies given by eq. (A.31). The contributions from distant longitudinal and transverse interactions are

$$\frac{\mathrm{d}\sigma_{\mathrm{dis,l}}}{\mathrm{d}W} = \frac{2\pi e^4}{\mathrm{m_e}v^2}\sum_k f_k \frac{1}{W_k} \ln\left(\frac{W_k}{Q_-}\frac{Q_- + 2\mathrm{m_e}c^2}{W_k + 2\mathrm{m_e}c^2}\right)\delta(W-W_k)\,\Theta(W_k-Q_-) \tag{3.79}$$

and

$$\frac{\mathrm{d}\sigma_{\mathrm{dis,t}}}{\mathrm{d}W} = \frac{2\pi e^4}{\mathrm{m_e}v^2}\sum_k f_k \frac{1}{W_k}\left\{\ln\left(\frac{1}{1-\beta^2}\right) - \beta^2 - \delta_{\mathrm{F}}\right\}\delta(W-W_k)\,\Theta(W_k-Q_-), \tag{3.80}$$

respectively. The energy-loss DCS for close collisions is

$$\frac{\mathrm{d}\sigma_{\mathrm{clo}}^{(\pm)}}{\mathrm{d}W} = \frac{2\pi e^4}{\mathrm{m_e}v^2}\sum_k f_k \frac{1}{W^2} F^{(\pm)}(E,W)\,\Theta(W-W_k). \tag{3.81}$$

The PDF of the energy loss in a single inelastic collision is given by

$$p_{\mathrm{in}}(W) = \frac{1}{\sigma_{\mathbf{in}}}\frac{\mathrm{d}\sigma_{\mathrm{in}}}{\mathrm{d}W}, \tag{3.82}$$

where

$$\sigma_{\mathrm{in}} = \int_0^{W_{\mathrm{max}}} \frac{\mathrm{d}\sigma_{\mathrm{in}}}{\mathrm{d}W}\,\mathrm{d}W \tag{3.83}$$

is the total cross section for inelastic interactions. It is convenient to introduce the quantities

$$\sigma_{\mathrm{in}}^{(n)} \equiv \int_0^{W_{\mathrm{max}}} W^n \frac{\mathrm{d}\sigma_{\mathrm{in}}}{\mathrm{d}W}\,\mathrm{d}W = \sigma_{\mathrm{in}}\int_0^{W_{\mathrm{max}}} W^n p_{\mathrm{in}}(W)\,\mathrm{d}W = \sigma_{\mathrm{in}}\langle W^n\rangle, \tag{3.84}$$

where $\langle W^n \rangle$ denotes the n-th moment of the energy loss in a single collision (notice that $\sigma_{\text{in}}^{(0)} = \sigma_{\text{in}}$). $\sigma_{\text{in}}^{(1)}$ and $\sigma_{\text{in}}^{(2)}$ are known as the stopping cross section and the energy straggling cross section (for inelastic collisions), respectively.

The mean free path λ_{in} for inelastic collisions is

$$\lambda_{\text{in}}^{-1} = \mathcal{N}\sigma_{\text{in}}, \tag{3.85}$$

where $\mathcal{N}$ is the number of scattering centres (atoms or molecules) per unit volume. The stopping power S_{in} and the energy straggling parameter Ω_{in}^2 are defined by

$$S_{\text{in}} = \mathcal{N}\sigma_{\text{in}}^{(1)} = \frac{\langle W \rangle}{\lambda_{\text{in}}} \tag{3.86}$$

and

$$\Omega_{\text{in}}^2 = \mathcal{N}\sigma_{\text{in}}^{(2)} = \frac{\langle W^2 \rangle}{\lambda_{\text{in}}}. \tag{3.87}$$

Notice that the stopping power gives the average energy loss per unit path length[4]. The physical meaning of the straggling parameter is less direct. Consider a monoenergetic electron (or positron) beam of energy E that impinges normally on a foil of material of (small) thickness $\mathrm{d}s$, and assume that the electrons do not scatter (i.e. they are not deflected) in the foil. The product $\Omega_{\text{in}}^2\,\mathrm{d}s$ then gives the variance of the energy distribution of the beam after traversing the foil (see also section 4.2).

The integrated cross sections $\sigma_{\text{in}}^{(n)}$ can be calculated as

$$\sigma_{\text{in}}^{(n)} = \sigma_{\text{dis,l}}^{(n)} + \sigma_{\text{dis,t}}^{(n)} + \sigma_{\text{clo}}^{(n)}. \tag{3.88}$$

The contributions from distant longitudinal and transverse interactions are

$$\sigma_{\text{dis,l}}^{(n)} = \frac{2\pi e^4}{\mathrm{m_e}v^2} \sum_k f_k W_k^{n-1} \ln\left(\frac{W_k}{Q_-}\frac{Q_- + 2\mathrm{m_e}c^2}{W_k + 2\mathrm{m_e}c^2}\right) \Theta(W_{\max} - W_k) \tag{3.89}$$

and

$$\sigma_{\text{dis,t}}^{(n)} = \frac{2\pi e^4}{\mathrm{m_e}v^2} \sum_k f_k W_k^{n-1} \left\{\ln\left(\frac{1}{1-\beta^2}\right) - \beta^2 - \delta_{\text{F}}\right\} \Theta(W_{\max} - W_k), \tag{3.90}$$

respectively. Notice that for distant interactions $W_{\max} = E$, for both electrons and positrons.

The integrated cross sections for close collisions are

$$\sigma_{\text{clo}}^{(n)} = \frac{2\pi e^4}{\mathrm{m_e}v^2} \sum_k f_k \int_{W_k}^{W_{\max}} W^{n-2} F^{(\pm)}(E, W)\,\mathrm{d}W. \tag{3.91}$$

In the case of electrons, the integrals in this formula are of the form

$$J_n^{(-)} = \int W^{n-2}\left[1 + \left(\frac{W}{E-W}\right)^2 - \frac{(1-a)W}{E-W} + \frac{aW^2}{E^2}\right] \mathrm{d}W \tag{3.92}$$

[4]The term "stopping power" is somewhat misleading; in fact, S_{in} has the dimensions of force.

and can be calculated analytically. For the orders 0, 1 and 2 we have

$$J_0^{(-)} = -\frac{1}{W} + \frac{1}{E-W} + \frac{1-a}{E}\ln\left(\frac{E-W}{W}\right) + \frac{aW}{E^2}, \tag{3.93}$$

$$J_1^{(-)} = \ln W + \frac{E}{E-W} + (2-a)\ln(E-W) + \frac{aW^2}{2E^2} \tag{3.94}$$

and

$$J_2^{(-)} = (2-a)W + \frac{2E^2-W^2}{E-W} + (3-a)E\ln(E-W) + \frac{aW^3}{3E^2}. \tag{3.95}$$

For positrons, the integrals in (3.91),

$$J_n^{(+)} \equiv \int W^{n-2}\left[1 - b_1\frac{W}{E} + b_2\left(\frac{W}{E}\right)^2 - b_3\left(\frac{W}{E}\right)^3 + b_4\left(\frac{W}{E}\right)^4\right] \mathrm{d}W, \tag{3.96}$$

can also be evaluated analytically as

$$J_0^{(+)} = -\frac{1}{W} - b_1\frac{\ln W}{E} + b_2\frac{W}{E^2} - b_3\frac{W^2}{2E^3} + b_4\frac{W^3}{3E^4}, \tag{3.97}$$

$$J_1^{(+)} = \ln W - b_1\frac{W}{E} + b_2\frac{W^2}{2E^2} - b_3\frac{W^3}{3E^3} + b_4\frac{W^4}{4E^4} \tag{3.98}$$

and

$$J_2^{(+)} = W - b_1\frac{W^2}{2E} + b_2\frac{W^3}{3E^2} - b_3\frac{W^4}{4E^3} + b_4\frac{W^5}{5E^4}. \tag{3.99}$$

Our analytical GOS model provides quite an accurate *average* description of inelastic collisions (see below). However, the continuous energy loss spectrum associated with single distant excitations of a given atomic electron shell is approximated here as a single resonance (a δ-distribution). As a consequence, the simulated energy loss spectra show unphysically narrow peaks at energy losses that are multiples of the resonance energies. These spurious peaks are automatically smoothed out after multiple inelastic collisions and also when the bin width used to tally the energy loss distributions is larger than the difference between resonance energies of neighbouring oscillators.

Fig. 3.9 displays total inelastic cross sections for electrons in aluminium and gold, as well as contributions from various groups of shells, as functions of the kinetic energy of the projectile. The curves labelled "K" and "L1+..." represent cross sections for ionization in these shells. The cross section for ionization in a bound shell decreases rapidly with the shell ionization energy U_i (since energy transfers less than U_i, which would promote the target electron to occupied states, are forbidden). As a consequence, collisions occur preferentially with electrons in the conduction band and in outer bound shells. Inner-shell ionization by electron/positron impact is a relatively unlikely process. It should be noted that our GOS model is too crude to provide an accurate description of inner-shell ionization. To illustrate this, fig. 3.9 includes K-shell ionization cross sections obtained from a more realistic semiempirical approximation (Mayol and Salvat, 1990), which agree reasonably well with experimental data. We see that there are significant

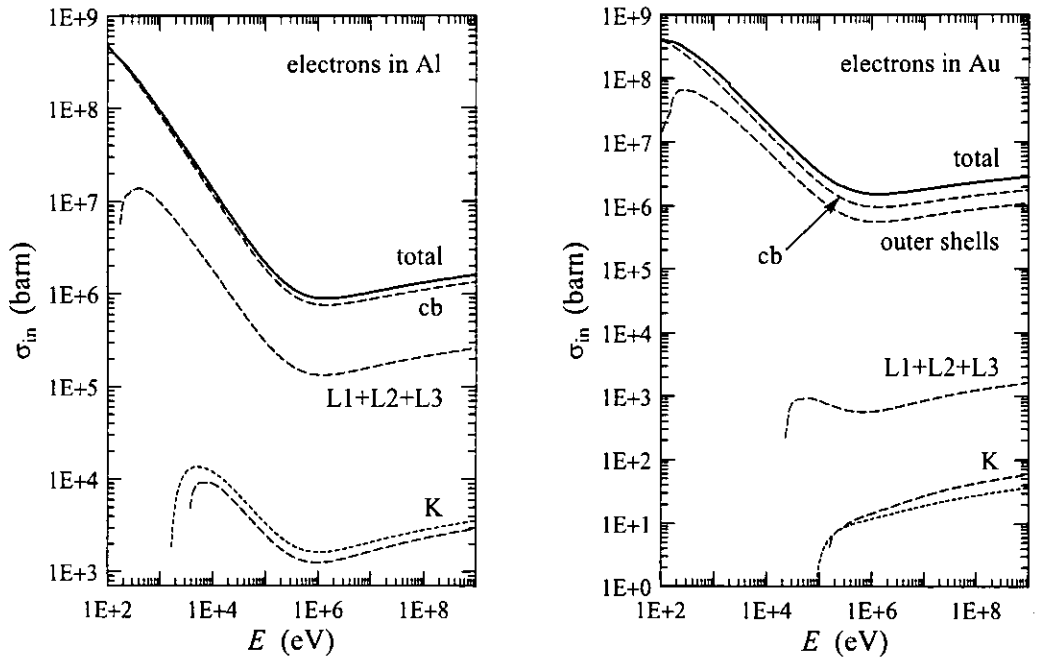


Figure 3.9: Total inelastic cross sections for electrons in aluminium and gold and contributions from the K-shell, L-shell, conduction band (cb) and outer shells, calculated from our model GOS ignoring density effect corrections (i.e. with $\delta_F = 0$). The short-dashed lines represent K-shell ionization cross sections calculated from a more elaborate theory (Mayol and Salvat, 1990), which yields results in close agreement with experimental data. Note: 1 barn=10^{-24} cm^2.

differences between the cross sections from the semiempirical approximation and the predictions of our model, which is designed to yield accurate stopping powers only. To get a realistic picture of inner-shell ionization, we should rely on much more elaborate physical schemes. In fact, even the Born approximation ceases to be appropriate for projectiles with kinetic energies near the ionization threshold.

Collision stopping powers for electrons in aluminium, silver and gold obtained from the present analytical model are compared with sample values from the ICRU37 (1984) stopping power tables [given also in Berger and Seltzer (1982)] for $E \geq 10$ keV in fig. 3.10. At these energies, our results practically coincide with the values in the tables of reference. In fig. 3.11, inelastic mean free paths and stopping powers for low-energy electrons ($E = 100$ eV to 100 keV) in aluminium and gold obtained from the present model are compared with experimental data from several authors. We see that the theory predicts the energy variation of total integrated cross sections down to relatively low energies. It should be noted that the adopted value of W_{cb}, the resonance energy of conduction band electrons, has a strong effect on the calculated mean free paths. In the case of free-electron-like materials such as aluminium, W_{cb} can be identified with

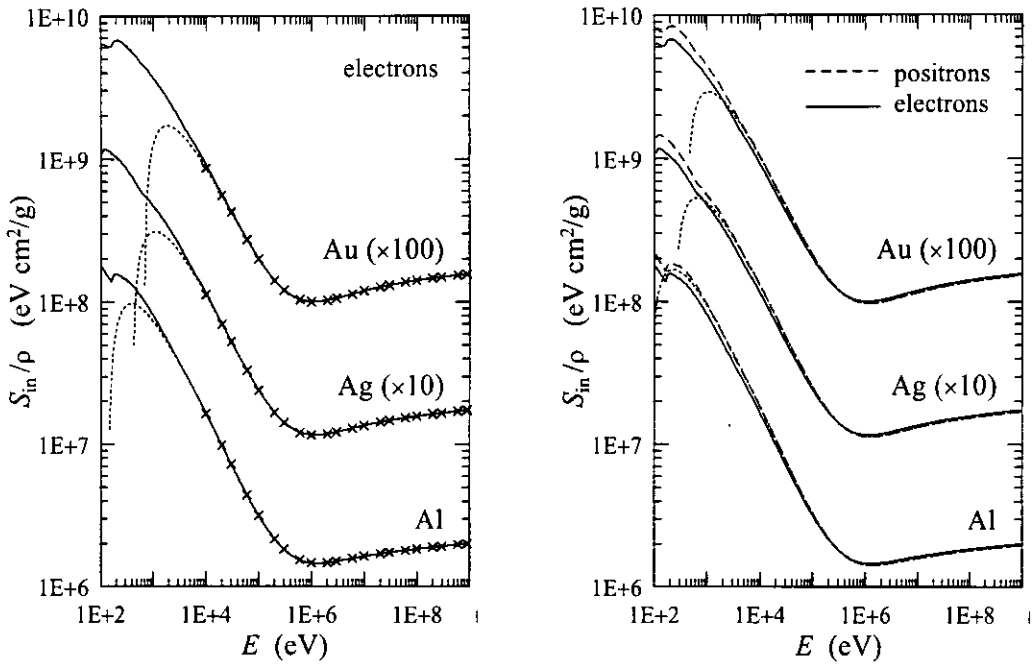


Figure 3.10: Collision stopping power S_{in}/ρ for electrons and positrons in Al, Ag (×10) and Au (×100) as a function of the kinetic energy. Continuous and dashed curves are results from the present model. Crosses are data from the ICRU37 tables (1984) [also, Berger and Seltzer, 1982)]. The dotted curves are predictions from the Bethe formula (3.105), for electrons and positrons.

the energy of plasmon excitations (which is the dominant energy-loss mechanism). For other solids, the outermost electrons have a broad energy loss spectrum and there is no simple way of predicting this parameter. Fortunately, the stopping power (and, hence, the global stopping process) is practically independent of the adopted value of W_{cb}. To generate the data for aluminium, Fig. 3.11, we have set $W_{cb} = 15$ eV, which is the measured energy of volume plasmons in the metal [eq. (3.51) with $f_{cb} = 3$ conduction electrons per atom gives $W_{cb} = 15.8$ eV]; in this case, the calculated mean free paths are seen to agree fairly well with measured data. In the case of gold, eq. (3.51) with $f_{cb} = 11$ conduction electrons per atom gives $W_{cb} = 30$ eV. Fig. 3.11 shows stopping powers and mean free paths for electrons in gold obtained with $W_{cb} = 30$ and 40 eV. We see that, as indicated above, the mean free path varies strongly with this parameter, but the stopping power is practically insensitive to it.

3.2.4 Stopping power of high-energy electrons and positrons

It is of interest to evaluate explicitly the stopping power for projectiles with high energies ($E \gg U_k$). We shall assume that $U_k \ll 2m_e c^2$ (for the most unfavourable case of the

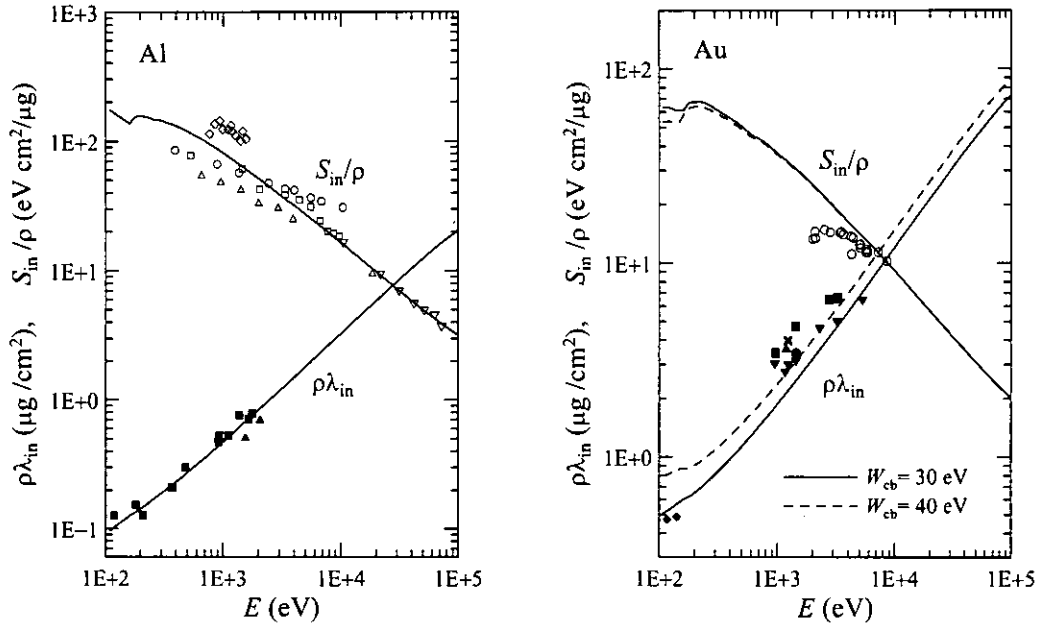


Figure 3.11: Collision mean free path and stopping power for low-energy electrons in Al and Au. The plotted quantities are $\rho\lambda_{in}$ and S_{in}/ρ. Special symbols are experimental data from different sources (see Fernández-Varea et al., 1993a); closed symbols for mean free paths and open symbols for stopping powers.

K shell of heavy elements, U_k is of the order of $2m_ec^2/10$). Under these circumstances, $Q_- \ll 2m_ec^2$ and we can use the approximation [see eq. (A.35)]

$$Q_- \simeq W_k^2/(2m_ec^2\beta^2). \tag{3.100}$$

The contribution from distant (longitudinal and transverse) interactions to the stopping cross section is then [see eqs. (3.79) and (3.80)]

$$\sigma_{\rm dis}^{(1)} \simeq \frac{2\pi e^4}{m_e v^2}\sum_k f_k\left\{\ln\left(\frac{2m_ec^2}{W_k}\right)+\ln\left(\frac{1}{1-\beta^2}\right)-\beta^2-\delta_{\rm F}\right\}. \tag{3.101}$$

The contribution of close interactions is given by

$$\sigma_{\rm clo}^{(1)} = \frac{2\pi e^4}{m_e v^2}\sum_k f_k\int_{W_k}^{W_{\rm max}} W^{-1}F^{(\pm)}(E,W)\,{\rm d}W. \tag{3.102}$$

Recalling that $E \gg U_k$, we have

$$\sigma_{\rm clo}^{(1)} \simeq \frac{2\pi e^4}{m_e v^2}\sum_k f_k\left\{\ln\left(\frac{E}{W_k}\right)+1-\left[1+\beta^2+2\sqrt{1-\beta^2}\right]\ln 2 + \frac{1}{8}\left(1-\sqrt{1-\beta^2}\right)^2\right\} \tag{3.103}$$

 3582

for electrons and

$$\sigma_{\rm clo}^{(1)} \simeq \frac{2\pi e^4}{{\rm m_e}v^2}\sum_k f_k \left\{ \ln\left(\frac{E}{W_k}\right) - b_1 + \frac{b_2}{2} - \frac{b_3}{3} + \frac{b_4}{4} \right\} \tag{3.104}$$

for positrons. Adding the distant and close stopping cross sections, and using the relation (3.56), we arrive at the familiar Bethe formula for the stopping power,

$$S_{\rm in} \equiv \mathcal{N}\left(\sigma_{\rm dis}^{(1)} + \sigma_{\rm clo}^{(1)}\right) = \mathcal{N}\frac{2\pi e^4}{{\rm m_e}v^2} Z \left\{ \ln\left(\frac{E^2}{I^2}\,\frac{\gamma+1}{2}\right) + f^{(\pm)}(\gamma) - \delta_{\rm F} \right\}, \tag{3.105}$$

where

$$f^{(-)}(\gamma) = 1 - \beta^2 - \frac{2\gamma-1}{\gamma^2}\ln 2 + \frac{1}{8}\left(\frac{\gamma-1}{\gamma}\right)^2 \tag{3.106}$$

and

$$f^{(+)}(\gamma) = 2\ln 2 - \frac{\beta^2}{12}\left[23 + \frac{14}{\gamma+1} + \frac{10}{(\gamma+1)^2} + \frac{4}{(\gamma+1)^3}\right] \tag{3.107}$$

for electrons and positrons, respectively. This formula can be derived from very general arguments that do not require knowing the fine details of the GOS; the only information needed is contained in the Bethe sum rule (3.45) and in the definition (3.46) of the mean excitation energy (see e.g. Fano, 1963). Since our approximate analytical GOS model is physically motivated, it satisfies the sum rule and reproduces the adopted value of the mean ionization energy, it yields the exact Bethe formula.

It is striking that the "asymptotic" Bethe formula is in fact valid down to fairly small energies, of the order of 10 keV for high-Z materials (see fig. 3.10). It also accounts for the differences between the stopping powers of electrons and positrons (to the same degree as our GOS model approximation).

For ultrarelativistic projectiles, for which the approximation (3.66) holds, the Bethe formula simplifies to

$$S_{\rm in} \simeq \mathcal{N}\frac{2\pi e^4}{{\rm m_e}v^2} Z \left\{ \ln\left(\frac{E^2}{\Omega_{\rm p}^2}\,\frac{\gamma+1}{2\gamma^2}\right) + f^{(\pm)}(\gamma) + 1 \right\}. \tag{3.108}$$

The mean excitation energy I has disappeared from this formula, showing that at very high energies the stopping power depends only on the electron density $\mathcal{N}Z$ of the medium.

3.2.5 Simulation of hard inelastic collisions

The DCSs given by expressions (3.78)-(3.81) permit the random sampling of the energy loss W and the angular deflection θ by using purely analytical methods. In the following we consider the case of mixed (class II) simulation, in which only hard collisions, with energy loss larger than a specified cutoff value $W_{\rm cc}$, are simulated (see chapter 4). As

the value of the cutoff energy loss can be selected arbitrarily, the sampling algorithm can also be used in detailed (interaction-by-interaction) simulations ($W_{cc} = 0$).

The first stage of the simulation is the selection of the active oscillator, for which we need to know the restricted total cross section,

$$\sigma(W_{cc}) = \int_{W_{cc}}^{W_{max}} \frac{d\sigma_{in}}{dW}\, dW = \sigma_{dis,l}(W_{cc}) + \sigma_{dis,t}(W_{cc}) + \sigma_{clo}(W_{cc})$$
$$= \sum_k \sigma_k(W_{cc}), \tag{3.109}$$

as well as the contribution of each oscillator, $\sigma_k(W_{cc})$. The active oscillator is sampled from the point probabilities $p_k = \sigma_k(W_{cc})/\sigma(W_{cc})$. Since these probabilities are calculated analytically, the sampling algorithm is relatively slow. In mixed simulations, the algorithm can be sped up by using a larger cutoff energy loss W_{cc}, which eliminates all the oscillators with $W_k < W_{cc}$ from the sum.

After selecting the active oscillator, the oscillator branch (distant or close) is determined and, finally, the variables W and Q (or $\cos\theta$) are sampled from the associated DCS. For close collisions, $Q = W$ and, therefore, the scattering angle is obtained directly from the energy loss.

Distant interactions

In distant interactions with the k-th oscillator, $W = W_k$. The contributions of transverse and longitudinal interactions to the restricted cross section define the relative probabilities of these interaction modes. If the interaction is (distant) transverse, the angular deflection of the projectile is neglected, i.e. $\cos\theta = 1$. For distant longitudinal collisions, the (unnormalized) PDF of Q is given by [see eq. (3.60)]

$$P_{dk}(Q) = \begin{cases} \dfrac{1}{Q\,[1 + Q/(2m_e c^2)]} & \text{if } Q_- < Q < W_k, \\ 0 & \text{otherwise,} \end{cases} \tag{3.110}$$

where Q_- is the minimum recoil energy, eq. (A.31). Random sampling from this PDF can be performed by the inverse transform method, which gives the sampling formula

$$Q = Q_S \left\{ \left[\frac{Q_S}{W_k} \left(1 + \frac{W_k}{2m_e c^2} \right) \right]^{\xi} - \frac{Q_S}{2m_e c^2} \right\}^{-1}, \tag{3.111}$$

where

$$Q_S \equiv \frac{Q_-}{1 + Q_-/(2m_e c^2)}. \tag{3.112}$$

Once the energy loss and the recoil energy have been sampled, the polar scattering angle θ is determined from eq. (A.40),

$$\cos\theta = \frac{E(E + 2m_e c^2) + (E - W)(E - W + 2m_e c^2) - Q(Q + 2m_e c^2)}{2\sqrt{E(E + 2m_e c^2)\,(E - W)(E - W + 2m_e c^2)}}. \tag{3.113}$$

 3534

The azimuthal scattering angle ϕ is sampled uniformly in the interval $(0, 2\pi)$.

Hard close collisions of electrons

For the formulation of the sampling algorithm, it is convenient to introduce the reduced energy loss $\kappa \equiv W/E$. The PDF of κ in close collisions of electrons with the k-th oscillator is given by [see eqs. (3.71) and (3.72)]

$$P_k^{(-)}(\kappa) \equiv \kappa^{-2} F^{(-)}(E, W)\, \Theta(\kappa - \kappa_c)\, \Theta\left(\frac{1}{2} - \kappa\right) = \left[\frac{1}{\kappa^2} + \frac{1}{(1-\kappa)^2} - \frac{1}{\kappa(1-\kappa)} + a\left(1 + \frac{1}{\kappa(1-\kappa)}\right)\right] \Theta(\kappa - \kappa_c)\, \Theta\left(\frac{1}{2} - \kappa\right), \qquad (3.114)$$

with $\kappa_c \equiv \max(W_k, W_{cc})/E$. Notice that the maximum allowed value of κ is $1/2$. Here, normalization is irrelevant.

We introduce the distribution

$$\Phi^{(-)}(\kappa) \equiv (\kappa^{-2} + 5a)\, \Theta(\kappa - \kappa_c)\, \Theta\left(\frac{1}{2} - \kappa\right), \qquad a \equiv \left(\frac{\gamma - 1}{\gamma}\right)^2. \qquad (3.115)$$

It may be shown that $\Phi^{(-)} > P_k^{(-)}$ in the interval $(\kappa_c, \frac{1}{2})$. Therefore, we can sample the reduced energy loss κ from the PDF (3.114) by using the rejection method (see section 1.2.4) with trial values sampled from the distribution (3.114) and acceptance probability $P_k^{(-)}/\Phi^{(-)}$.

Random sampling from the PDF (3.115), can be performed by using the composition method (section 1.2.5). We consider the following decomposition of the (normalized) PDF given by eq. (3.115):

$$\Phi^{(-)}_{\rm norm}(\kappa) = \frac{1}{1 + 5a\kappa_c/2}\left[p_1(\kappa) + (5a\kappa_c/2) p_2(\kappa)\right], \qquad (3.116)$$

where

$$p_1(\kappa) = \frac{\kappa_c}{1 - 2\kappa_c}\kappa^{-2}, \qquad p_2(\kappa) = \frac{2}{1 - 2\kappa_c} \qquad (3.117)$$

are normalized PDFs in the interval $(\kappa_c, \frac{1}{2})$. Random values of κ from the PDF (3.115) can be generated by using the following algorithm:

(i) Generate ξ.

(ii) Set $\zeta = (1 + 5a\kappa_c/2)\xi$.

(iii) If $\zeta < 1$, deliver the value $\kappa = \kappa_c/[1 - \zeta(1 - 2\kappa_c)]$.

(iv) If $\zeta > 1$, deliver the value $\kappa = \kappa_c + (\zeta - 1)(1 - 2\kappa_c)/(5a\kappa_c)$.

The rejection algorithm for random sampling of κ from the PDF (3.114) proceeds as follows:

(i) Sample κ from the distribution given by eq. (3.115).

(ii) Generate a random number ξ.

(iii) If $\xi(1+5a\kappa^2) < \kappa^2 P_k^{(-)}(\kappa)$, deliver κ.

(iv) Go to step (i).

Notice that in the third step we accept the κ value with probability $P_k^{(-)}/\Phi^{(-)}$, which approaches unity when κ is small.

The efficiency of this sampling method depends on the values of the energy E and the cutoff reduced energy loss κ_c, as shown in table 3.1. For a given energy and for W_{cc} values which are not too large, the efficiency increases when W_{cc} decreases.

Table 3.1: Efficiency (%) of the random sampling algorithm of the energy loss in close collisions of electrons and positrons for different values of the energy E and the cutoff energy loss κ_c.

E (eV)	κ_c				
	0.001	0.01	0.1	0.25	0.4
10^3	99.9	99.9	99.8	99.7	99.6
10^5	99.7	98	87	77	70
10^7	99	93	70	59	59
10^9	99	93	71	62	63

After sampling the energy loss $W = \kappa E$, the polar scattering angle θ is obtained from eq. (A.40) with $Q = W$. This yields

$$\cos^2\theta = \frac{E-W}{E}\,\frac{E+2\mathrm{m_e}c^2}{E-W+2\mathrm{m_e}c^2}, \tag{3.118}$$

which agrees with eq. (A.17). The azimuthal scattering angle ϕ is sampled uniformly in the interval $(0, 2\pi)$.

Hard close collisions of positrons

The PDF of the reduced energy loss $\kappa \equiv W/E$ in positron close collisions with the k-th oscillator is given by [see eqs. (3.76) and (3.77)]

$$\begin{aligned} P_k^{(+)}(\kappa) &= \kappa^{-2} F_k^{(+)}(E,W)\,\Theta(\kappa-\kappa_c)\,\Theta(1-\kappa) \\ &= \left[\frac{1}{\kappa^2} - \frac{b_1}{\kappa} + b_2 - b_3\kappa + b_4\kappa^2\right]\Theta(\kappa-\kappa_c)\,\Theta(1-\kappa) \end{aligned} \tag{3.119}$$

with $\kappa_c \equiv \max(W_k, W_{cc})/E$. The maximum allowed reduced energy loss is 1. Again, normalization is not important.

Consider the distribution

$$\Phi^{(+)}(\kappa) \equiv \kappa^{-2}\Theta(\kappa - \kappa_c)\,\Theta(1-\kappa). \tag{3.120}$$

It is easy to see that $\Phi^{(+)} > P_k^{(+)}$ in the interval $(\kappa_c, 1)$. Therefore, we can generate κ from the PDF, eq. (3.119), by using the rejection method with trial values sampled from the distribution of eq. (3.120) and acceptance probability $P_k^{(+)}/\Phi^{(+)}$. Sampling from the PDF $\Phi^{(+)}$ can easily be performed with the inverse transform method.

The algorithm for random sampling from the PDF (3.119), is:

(i) Sample κ from the PDF (3.120), as $\kappa = \kappa_c/[1 - \xi(1-\kappa_c)]$.

(ii) Generate a new random number ξ.

(iii) If $\xi < \kappa^2 P_k^{(+)}(\kappa)$, deliver κ.

(iv) Go to step (i).

The efficiency of this algorithm, for given values of the kinetic energy and the cutoff reduced energy loss κ_c, practically coincides with that of the algorithm for electron collisions described above (see table 3.1).

Secondary electron emission

According to our GOS model, each oscillator W_k corresponds to a shell with f_k electrons and ionization energy U_k. As discussed above (see section 3.2.2), in the case of ionization of an inner shell i we consider that a secondary electron (delta ray) is emitted with energy $E_s = W - U_i$ and that the residual ion is left with a vacancy in the shell i. In the case of ionization of outer shells, the simulated delta ray is emitted with kinetic energy $E_s = W$ and the target atom is assumed to remain in its ground state. To set the initial direction of the delta ray, we assume that the target electron was initially at rest, i.e. the delta ray is emitted in the direction of the momentum transfer $\mathbf{q}$. This implies that the polar emission angle θ_s (see fig. 3.1) coincides with the recoil angle θ_r [which is given by eq. (A.42)],

$$\cos^2\theta_s = \frac{W^2/\beta^2}{Q(Q+2m_ec^2)}\left(1 + \frac{Q(Q+2m_ec^2) - W^2}{2W(E+m_ec^2)}\right)^2. \tag{3.121}$$

In the case of close collisions ($Q = W$), this expression simplifies to

$$\cos\theta_s\,(Q=W) = \left(\frac{W}{E}\,\frac{E+2m_ec^2}{W+2m_ec^2}\right)^{1/2}, \tag{3.122}$$

which agrees with the result for binary collisions with free electrons at rest, see eq. (A.18). Since the momentum transfer lies on the scattering plane (i.e. on the plane

formed by the initial and final momenta of the projectile), the azimuthal emission angle is $\phi_s = \pi + \phi$.

In reality, the target electrons are not at rest and, therefore, the angular distribution of emitted delta rays is broad. Since the average momentum of bound electrons is zero, the average direction of delta rays coincides with the direction of $\mathbf{q}$. Thus, our simple emission model correctly predicts the average initial direction of delta rays, but disregards the "Doppler broadening" of the angular distribution. This is not a serious drawback, because secondary electrons are usually emitted with initial kinetic energies that are much smaller than the initial energy of the projectile. This means that the direction of motion of the delta ray is randomized, by elastic and inelastic collisions, after a relatively short path length (much shorter than the transport mean free path of the projectile).

3.3 Bremsstrahlung emission

As a result of the acceleration caused by the electrostatic field of atoms, swift electrons (or positrons) emit bremsstrahlung (braking radiation). In each bremsstrahlung event, an electron with kinetic energy E generates a photon of energy W, which takes values in the interval from 0 to E. The process is described by an atomic DCS, differential in the energy loss W, the final direction of the projectile and the direction of the emitted photon (Koch and Motz, 1959; Tsai, 1974). The habitual practice in Monte Carlo simulation is to sample the energy loss from the single-variable distribution obtained by integrating the DCS over the other variables. This permits the generation of W easily, but information on the angular distributions is completely lost and has to be regained from suitable approximations. Angular deflections of the projectile are considered to be accounted for by the elastic scattering DCS and, consequently, the direction of movement of the projectile is kept unaltered in the simulation of radiative events.

3.3.1 The energy-loss scaled DCS

A simple description of the bremsstrahlung DCS is provided by the Bethe-Heitler formula with screening, which is derived within the Born approximation (Bethe and Heitler, 1934; Tsai, 1974). Although this formula is valid only when the kinetic energy of the electron before and after photon emission is much larger than its rest energy $m_e c^2$, it accounts for the most relevant features of the emission process. Within the Born approximation, bremsstrahlung emission is closely related to electron-positron pair production. In particular, the Bethe-Heitler DCS formulae for pair production and bremsstrahlung emission involve the same screening functions. Considering the exponential screening model (2.78), the Bethe-Heitler DCS for bremsstrahlung emission by electrons in the field of an atom of atomic number Z and screening radius R can be expressed as (Salvat

and Fernández-Varea, 1992)

$$\frac{\mathrm{d}\sigma_{\mathrm{br}}^{(\mathrm{BH})}}{\mathrm{d}W} = r_{\mathrm{e}}^2 \alpha Z(Z+\eta)\frac{1}{W}\left[\epsilon^2\,\varphi_1(b) + \frac{4}{3}(1-\epsilon)\,\varphi_2(b)\right], \tag{3.123}$$

where α is the fine-structure constant, r_{e} is the classical electron radius,

$$\epsilon = \frac{W}{E + \mathrm{m_e}c^2} = \frac{W}{\gamma\,\mathrm{m_e}c^2}, \qquad b = \frac{R\mathrm{m_e}c}{\hbar}\,\frac{1}{2\gamma}\,\frac{\epsilon}{1-\epsilon}, \tag{3.124}$$

and

$$\begin{aligned}
\varphi_1(b) &= 4\ln(R\mathrm{m_e}c/\hbar) + 2 - 2\ln(1+b^2) - 4b\arctan(b^{-1}),\\
\varphi_2(b) &= 4\ln(R\mathrm{m_e}c/\hbar) + \frac{7}{3} - 2\ln(1+b^2) - 6b\arctan(b^{-1})\\
&\quad - b^2\left[4 - 4b\arctan(b^{-1}) - 3\ln(1+b^{-2})\right].
\end{aligned} \tag{3.125}$$

The quantity η in eq. (3.123) accounts for the production of bremsstrahlung in the field of the atomic electrons (see e.g. Seltzer and Berger, 1985); in the high-energy limit $\eta \simeq 1.2$.

The Bethe-Heitler formula indicates that, for a given value of Z, the quantity $W\mathrm{d}\sigma_{\mathrm{br}}/\mathrm{d}W$ varies smoothly with E and W. It is therefore customary to express the DCS for bremsstrahlung emission by electrons in the form

$$\frac{\mathrm{d}\sigma_{\mathrm{br}}}{\mathrm{d}W} = \frac{Z^2}{\beta^2}\,\frac{1}{W}\,\chi(Z,E,\kappa), \tag{3.126}$$

where W is the energy of the emitted photon, κ is the reduced photon energy, defined as

$$\kappa \equiv W/E, \tag{3.127}$$

which takes values between 0 and 1. The quantity

$$\chi(Z,E,\kappa) = (\beta^2/Z^2)W\frac{\mathrm{d}\sigma_{\mathrm{br}}}{\mathrm{d}W} \tag{3.128}$$

is known as the "scaled" bremsstrahlung DCS; for a given element Z, it varies smoothly with E and κ. Seltzer and Berger (1985, 1986) produced extensive tables of the scaled DCS for all the elements (Z =1–92) and for electron energies from 1 keV to 10 GeV. They tabulated the scaled DCSs for emission in the (screened) field of the nucleus (electron-nucleus bremsstrahlung) and in the field of atomic electrons (electron-electron bremsstrahlung) separately, as well as their sum, the total scaled DCS. The electron-nucleus bremsstrahlung DCS was calculated by combining analytical high-energy theories with results from partial-wave calculations by Pratt et al. (1977) for bremsstrahlung emission in screened atomic fields and energies below 2 MeV. The scaled DCS for

electron-electron bremsstrahlung was obtained from the theory of Haug (1975) combined with a screening correction that involves Hartree-Fock incoherent scattering functions. Seltzer and Berger's scaled DCS tables constitute the most reliable theoretical representation of bremsstrahlung energy spectra available at present.

The PENELOPE database of scaled bremsstrahlung DCSs consists of 92 files, one for each element from hydrogen to uranium, which were generated from the original database of Seltzer and Berger. The file of the element Z contains the values of $\chi(Z, E_i, \kappa_j)$ for a set of electron kinetic energies E_i, which covers the range from 1 keV to 10 GeV and is suitably spaced to allow accurate natural cubic spline interpolation in $\ln E$. For each energy E_i in this grid, the table contains the values of the scaled DCS for a given set of 32 reduced photon energies κ_j (the same for all elements), which span the interval (0,1), with a higher density at the upper end of this interval to reproduce the structure of the bremsstrahlung "tip" (see fig. 3.12). The spacing of the κ-grid is dense enough to allow linear interpolation of $\chi(Z, E_i, \kappa_j)$ in κ.

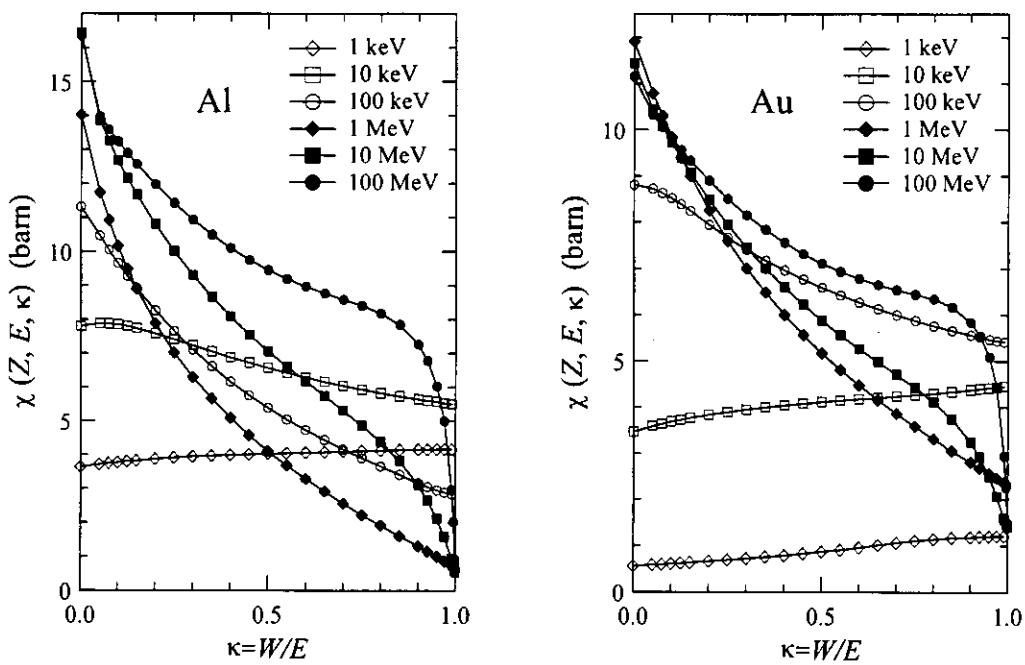


Figure 3.12: Numerical scaled bremsstrahlung energy-loss DCSs of Al and Au for electrons with the indicated energies (Seltzer and Berger, 1986).

In the case of compounds (or mixtures) we use the additivity rule and compute the molecular DCS as the sum of the DCSs of all the atoms in a molecule. Consider a compound X_xY_y, whose molecules consist of x atoms of the element X and y atoms of the element Y. The molecular DCS is

$$\frac{d\sigma_{br,mol}}{dW} = x\frac{Z_X^2}{\beta^2}\frac{1}{W}\chi(Z_X, E, \kappa) + y\frac{Z_Y^2}{\beta^2}\frac{1}{W}\chi(Z_Y, E, \kappa). \qquad (3.129)$$

3540

To simulate each radiative event in a compound, we should first select the element (X or Y) where the emission occurs and then sample the photon energy and direction from the corresponding atomic DCS. This is a lengthy process and requires storing the scaled DCSs for all the elements present. To simplify the simulation, we shall express the molecular DCS in the same form as the atomic DCS, eq. (3.126),

$$\frac{\mathrm{d}\sigma_{\mathrm{br,mol}}}{\mathrm{d}W} = \frac{Z_{\mathrm{eq}}^2}{\beta^2}\frac{1}{W}\chi_{\mathrm{mol}}(Z_{\mathrm{eq}}, E, \kappa), \tag{3.130}$$

where

$$Z_{\mathrm{eq}}^2 \equiv \frac{1}{\mathrm{x}+\mathrm{y}}\left(\mathrm{x}Z_{\mathrm{X}}^2 + \mathrm{y}Z_{\mathrm{Y}}^2\right) \tag{3.131}$$

is the "equivalent" atomic number Z_{eq} and

$$\chi_{\mathrm{mol}}(Z_{\mathrm{eq}}, E, \kappa) = \frac{\mathrm{x}Z_{\mathrm{X}}^2}{Z_{\mathrm{eq}}^2}\chi(Z_{\mathrm{X}}, E, \kappa) + \frac{\mathrm{y}Z_{\mathrm{Y}}^2}{Z_{\mathrm{eq}}^2}\chi(Z_{\mathrm{Y}}, E, \kappa) \tag{3.132}$$

is the molecular scaled DCS. Radiative events will be sampled directly from the molecular DCS (3.130). This method may introduce slight inconsistencies in the angular distribution of the emitted photons (see below), which usually have a negligible effect on the simulation results.

The radiative DCS for positrons reduces to that of electrons in the high-energy limit but is smaller for intermediate and low energies. Owing to the lack of more accurate calculations, the DCS for positrons is obtained by multiplying the electron DCS by a κ-independent factor, i.e.

$$\frac{\mathrm{d}\sigma_{\mathrm{br}}^{(+)}}{\mathrm{d}W} = F_{\mathrm{p}}(Z, E)\frac{\mathrm{d}\sigma_{\mathrm{br}}^{(-)}}{\mathrm{d}W}. \tag{3.133}$$

The factor $F_{\mathrm{p}}(Z, E)$ is set equal to the ratio of the radiative stopping powers for positrons and electrons, which has been calculated by Kim et al. (1986) (cf. Berger and Seltzer, 1982). In the calculations we use the following analytical approximation

$$\begin{aligned} F_{\mathrm{p}}(Z, E) = 1 - \exp(&-1.2359\times10^{-1}\,t + 6.1274\times10^{-2}\,t^2 - 3.1516\times10^{-2}\,t^3 \\ &+ 7.7446\times10^{-3}\,t^4 - 1.0595\times10^{-3}\,t^5 + 7.0568\times10^{-5}\,t^6 \\ &- 1.8080\times10^{-6}\,t^7), \end{aligned} \tag{3.134}$$

where

$$t = \ln\left(1 + \frac{10^6}{Z^2}\frac{E}{\mathrm{m_e}c^2}\right). \tag{3.135}$$

Expression (3.134) reproduces the values of $F_{\mathrm{p}}(Z, E)$ tabulated by Kim et al. (1986) to an accuracy of about 0.5%.

3.3.2 Integrated cross sections

The total cross section for bremsstrahlung emission is infinite due to the divergence of the DCS (3.126) for small reduced photon energies. Nevertheless, the cross section for emission of photons with reduced energy larger than a given cutoff value $W_{\rm cr}$ is finite. The corresponding mean free path is

$$\lambda_{\rm br}^{-1}(E;W_{\rm cr}) \equiv \mathcal{N}\int_{W_{\rm cr}}^{E}\frac{{\rm d}\sigma_{\rm br}}{{\rm d}W}\,{\rm d}W = \mathcal{N}\frac{Z^2}{\beta^2}\int_{\kappa_{\rm cr}}^{1}\frac{1}{\kappa}\,\chi(Z,E,\kappa)\,{\rm d}\kappa, \tag{3.136}$$

where $\kappa_{\rm cr} = W_{\rm cr}/E$. The radiative stopping power and the radiative energy straggling parameter, defined by

$$S_{\rm br}(E) \equiv \mathcal{N}\int_{0}^{E} W\frac{{\rm d}\sigma_{\rm br}}{{\rm d}W}\,{\rm d}W = \mathcal{N}\frac{Z^2}{\beta^2}E\int_{0}^{1}\chi(Z,E,\kappa)\,{\rm d}\kappa \tag{3.137}$$

and

$$\Omega_{\rm br}^2(E) \equiv \mathcal{N}\int_{0}^{E} W^2\frac{{\rm d}\sigma_{\rm br}}{{\rm d}W}\,{\rm d}W = \mathcal{N}\frac{Z^2}{\beta^2}E^2\int_{0}^{1}\kappa\,\chi(Z,E,\kappa)\,{\rm d}\kappa, \tag{3.138}$$

are both finite. For the kinetic energies E_i of the grid, these quantities are easily calculated from the tabulated scaled DCS by using linear interpolation in κ. For positrons, the definitions (3.136)-(3.138) must be multiplied by the factor $F_{\rm p}(Z,E)$ [eq. (3.134)].

Radiative stopping powers of aluminium, silver and gold for electrons and positrons are shown as functions of the kinetic energy in fig. 3.13. The stopping powers computed from the DCS given by eq. (3.126) practically coincide with ICRU37 (1984) values (also Berger and Seltzer, 1982). To leave room for future improvements, PENELOPE reads the radiative stopping power for electrons from the input material data file, and renormalizes the DCS, eq. (3.126), (i.e. multiplies it by a κ-independent factor) so as to exactly reproduce the input radiative stopping power.

CSDA range

As mentioned above, the stopping power gives the average energy loss per unit path length. Thus, when an electron/positron with kinetic energy E advances a small distance ${\rm d}s$ within a medium, it loses an (average) energy ${\rm d}E = -S(E){\rm d}s$, where

$$S(E) = S_{\rm in}(E) + S_{\rm br}(E) = -\frac{{\rm d}E}{{\rm d}s} \tag{3.139}$$

is the total (collisional+radiative) stopping power. Many electron transport calculations and old Monte Carlo simulations are based on the so-called continuous slowing down approximation (CSDA), which assumes that particles lose energy in a continuous way and at a rate equal to the stopping power. Evidently, the CSDA disregards energy-loss fluctuations and, therefore, it should be used with caution.

 3542

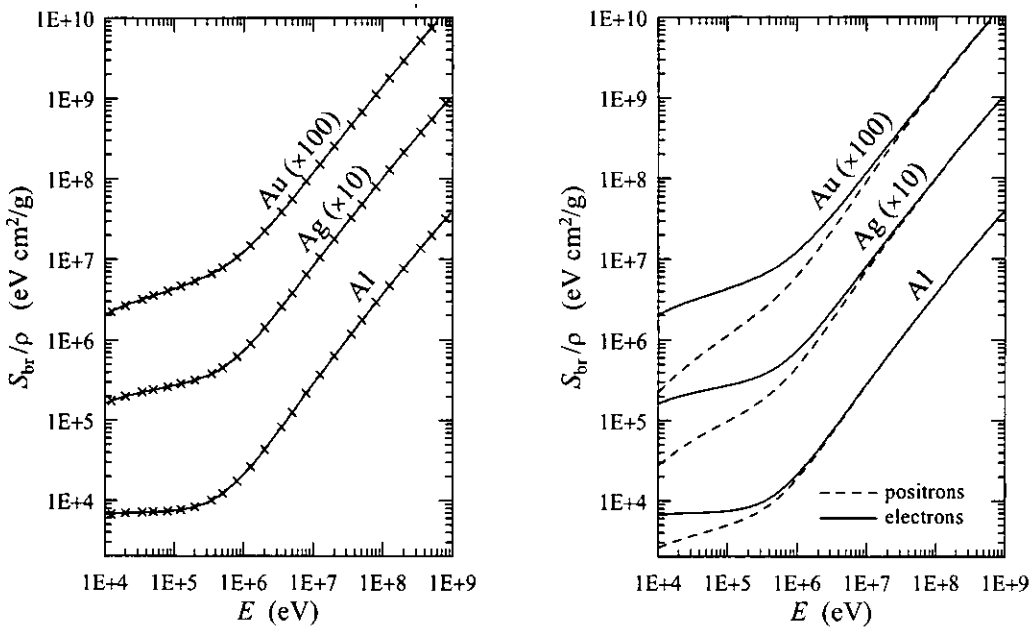


Figure 3.13: Radiative stopping power S_{br}/ρ for electrons and positrons in Al, Ag (×10) and Au (×100) as a function of the kinetic energy. Solid and dashed curves are results from the present model. Crosses are data from the ICRU37 report (1984) (also in Berger and Seltzer, 1982).

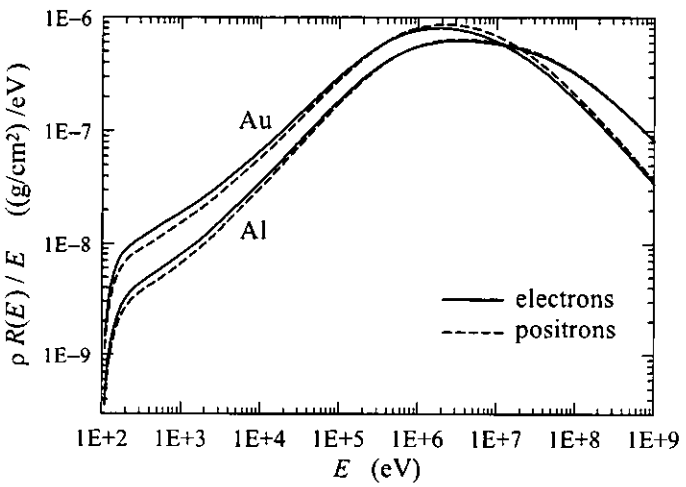


Figure 3.14: CSDA ranges for electrons and positrons in Al and Au as functions of the kinetic energy of the particle.

A parameter of practical importance is the so-called CSDA range (or Bethe range), which is defined as the path length travelled by a particle (in an infinite medium) before being absorbed and is given by

$$R(E) = \int_{E_{\rm abs}}^{E} \frac{\mathrm{d}E'}{S(E')}, \tag{3.140}$$

where we have considered that particles are effectively absorbed when they reach the energy $E_{\rm abs}$. Notice that the CSDA range gives the *average* path length, actual (or Monte Carlo generated) path lengths fluctuate about the mean $R(E)$; the distribution of ranges has been studied by Lewis (1952). Fig. 3.14 displays CSDA ranges for electrons and positrons in aluminium and gold, this information is useful e.g. in estimating the maximum penetration depth of a beam and for range rejection (a variance reduction method). Compare fig. 3.14 with figs. 3.10 and 3.13 (right plots only) to get a feeling of how differences in stopping power between electrons and positrons are reflected on the CSDA ranges of these particles.

3.3.3 Angular distribution of emitted photons

The direction of the emitted bremsstrahlung photon is determined by the polar angle θ (see fig. 3.1) and the azimuthal angle ϕ. For isotropic media, with randomly oriented atoms or molecules, the bremsstrahlung DCS is independent of ϕ and can be expressed as

$$\frac{\mathrm{d}^2\sigma_{\rm br}}{\mathrm{d}W\,\mathrm{d}(\cos\theta)} = \frac{\mathrm{d}\sigma_{\rm br}}{\mathrm{d}W}\,p(Z,E,\kappa;\cos\theta) = \frac{Z^2}{\beta^2}\,\frac{1}{W}\,\chi(Z,E,\kappa)\,p(Z,E,\kappa;\cos\theta), \tag{3.141}$$

where $p(Z,E,\kappa;\cos\theta)$ is the PDF of $\cos\theta$, θ is the polar angle of the photon direction relative to the direction of the projectile (fig. 3.1).

Numerical values of the "shape function" $p(Z,E,\kappa;\cos\theta)$, calculated by partial-wave methods, have been published by Kissel et al. (1983) for the following benchmark cases: Z =2, 8, 13, 47, 79, 92; E = 1, 5, 10, 50, 100, 500 keV and κ = 0, 0.6, 0.8, 0.95. These authors also gave a parameterization of the shape function in terms of Legendre polynomials. Unfortunately, their analytical form is not suited for random sampling of the photon direction. In PENELOPE we use a different parameterization that allows the random sampling of $\cos\theta$ in a simple way. Owing to the lack of numerical data for positrons, it is assumed that the shape function for positrons is the same as for electrons.

In previous simulation studies of x-ray emission from solids bombarded by electron beams (Acosta et al., 1998), the angular distribution of bremsstrahlung photons was described by means of the semiempirical analytical formulae derived by Kirkpatrick and Wiedmann (1945) [and subsequently modified by Statham (1976)]. These formulae were obtained by fitting the bremsstrahlung DCS derived from Sommerfeld's theory. The shape function obtained from the Kirkpatrick-Wiedmann-Statham fit reads

$$p^{(\rm KWS)}(Z,E,\kappa;\cos\theta) = \frac{\sigma_x(1-\cos^2\theta)+\sigma_y(1+\cos^2\theta)}{(1-\beta\cos\theta)^2}, \tag{3.142}$$

3544

where the quantities σ_x and σ_y are independent of θ. Although this simple formula predicts the global trends of the partial-wave shape functions of Kissel et al. (1983) in certain energy and atomic number ranges, its accuracy is not sufficient for general-purpose simulations. In a preliminary analysis, we tried to improve this formula and determined the parameters σ_x and σ_y by direct fitting to the numerical partial-wave shape functions, but the improvement was not substantial. However, this analysis confirmed that the analytical form (3.142) is flexible enough to approximate the "true" (partial-wave) shape.

The analytical form (3.142) is plausible even for projectiles with relatively high energies, say E larger than 1 MeV, for which the angular distribution of emitted photons is peaked at forward directions. This can be understood by means of the following classical argument (see e.g. Jackson, 1975). Assume that the incident electron is moving in the direction of the z-axis of a reference frame K at rest with respect to the laboratory frame. Let (θ', ϕ') denote the polar and azimuthal angles of the direction of the emitted photon in a reference frame K′ that moves with the electron and whose axes are parallel to those of K. In K′, we expect that the angular distribution of the emitted photons will not depart much from the isotropic distribution. To be more specific, we consider the following ansatz (modified dipole distribution) for the shape function in K′,

$$p_{\mathrm{d}}(\cos\theta') = A\frac{3}{8}(1+\cos^2\theta') + (1-A)\frac{3}{4}(1-\cos^2\theta'), \qquad (0 \le A \le 1), \tag{3.143}$$

which is motivated by the relative success of the Kirkpatrick-Wiedmann-Statham formula at low energies (note that the projectile is at rest in K′). The direction of emission (θ, ϕ) in K is obtained by means of the Lorentz transformation

$$\cos\theta = \frac{\cos\theta' + \beta}{1 + \beta\cos\theta'}, \qquad \phi = \phi'. \tag{3.144}$$

Thus, the angular distribution in K reads

$$\begin{aligned} p(\cos\theta) &= p_{\mathrm{d}}(\cos\theta')\,\frac{\mathrm{d}(\cos\theta')}{\mathrm{d}(\cos\theta)} \\ &= A\frac{3}{8}\left[1+\left(\frac{\cos\theta-\beta}{1-\beta\cos\theta}\right)^2\right]\frac{1-\beta^2}{(1-\beta\cos\theta)^2} \\ &\quad + (1-A)\frac{3}{4}\left[1-\left(\frac{\cos\theta-\beta}{1-\beta\cos\theta}\right)^2\right]\frac{1-\beta^2}{(1-\beta\cos\theta)^2}. \end{aligned} \tag{3.145}$$

Now, it is clear that when β tends to unity, the shape function concentrates at forward directions.

We found that the benchmark partial-wave shape functions of Kissel et al. (1983) can be closely approximated by the analytical form (3.145) if one considers A and β as

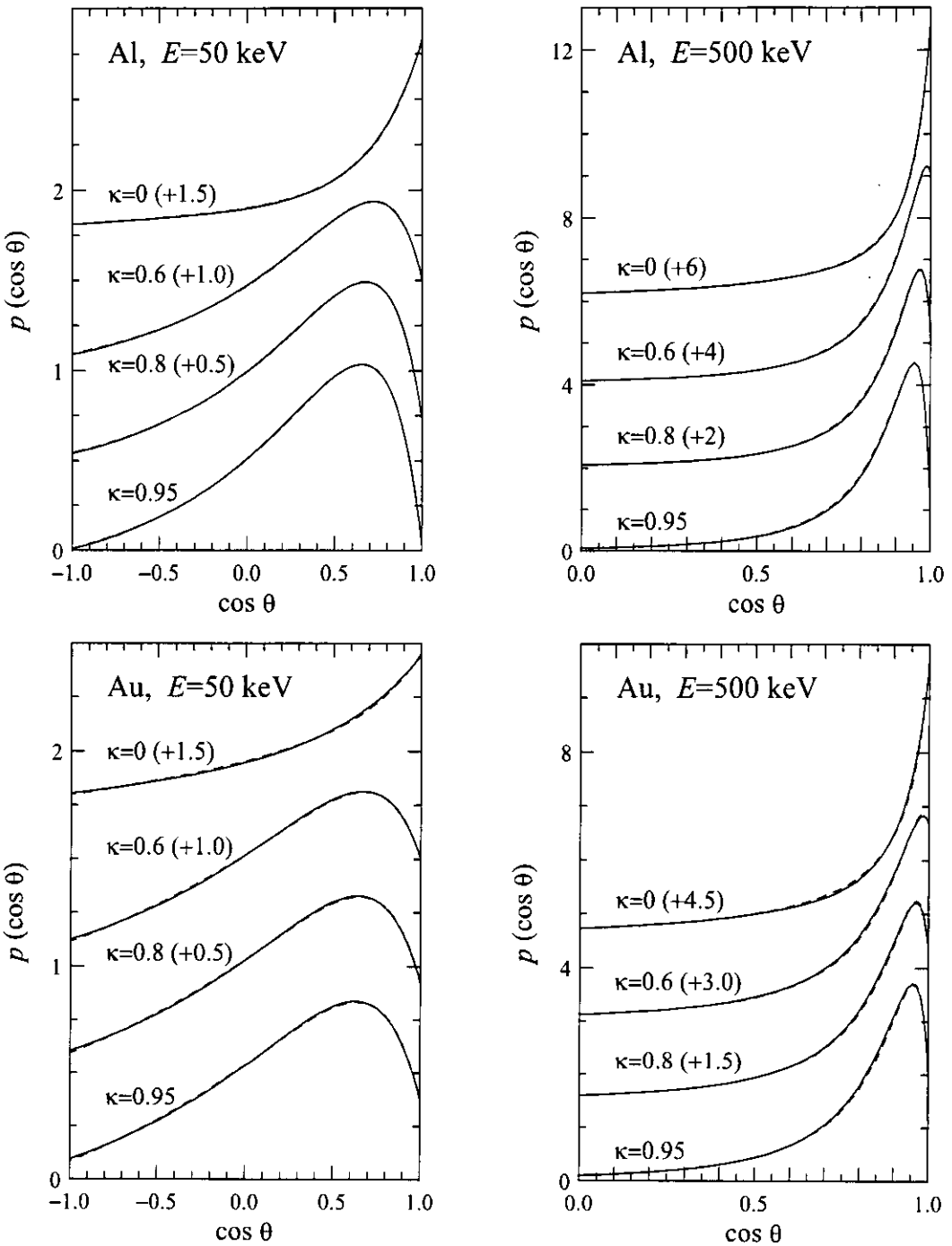


Figure 3.15: Shape functions (angular distributions) for bremsstrahlung emission by electrons of the indicated energies in the fields of Al and Au atoms. Dashed curves are partial-wave shape functions of Kissel et al. (1983). Continuous curves are the present analytical fits, eq. (3.146). For visual aid, some curves have been shifted upwards in the amounts indicated in parentheses.

 3524

adjustable parameters. Explicitly, we write

$$p_{\rm fit}(\cos\theta) = A\frac{3}{8}\left[1+\left(\frac{\cos\theta-\beta'}{1-\beta'\cos\theta}\right)^2\right]\frac{1-\beta'^2}{(1-\beta'\cos\theta)^2} + (1-A)\frac{3}{4}\left[1-\left(\frac{\cos\theta-\beta'}{1-\beta'\cos\theta}\right)^2\right]\frac{1-\beta'^2}{(1-\beta'\cos\theta)^2}, \qquad (3.146)$$

with $\beta' = \beta(1+B)$. The parameters A and B have been determined, by least squares fitting, for the 144 combinations of atomic number, electron energy and reduced photon energy corresponding to the benchmark shape functions tabulated by Kissel et al. (1983). Results of this fit are compared with the original partial-wave shape functions in fig. 3.15. The largest differences between the fits and the data were found for the higher atomic numbers, but even then the fits are very accurate, as shown in fig. 3.15. The quantities $\ln(AZ\beta)$ and $B\beta$ vary smoothly with Z, β and κ and can be obtained by cubic spline interpolation of their values for the benchmark cases. This permits the fast evaluation of the shape function for any combination of Z, β and κ. Moreover, the random sampling of the photon direction, i.e. of $\cos\theta$, can be performed by means of a simple, fast analytical algorithm (see below).

3.3.4 Simulation of hard radiative events

Let us now consider the simulation of hard radiative events ($W > W_{\rm cr}$) from the DCS defined by eqs. (3.141) and (3.146). PENELOPE reads the scaled bremsstrahlung DCS from the database files and, by natural cubic spline interpolation/extrapolation in $\ln E$, produces a table for a denser logarithmic grid of 200 energies (and for the "standard" mesh of 32 κ's), which is stored in memory. This energy grid spans the full energy range considered in the simulation and allows accurate (and fast) linear interpolation of the scaled DCS in the variable $\ln E$, which is more adequate than E when interpolation over a wide energy interval is required.

Notice that in the Monte Carlo simulation the kinetic energy of the transported electron (or positron) varies in a random way and may take arbitrary values within a certain domain. Hence, we must be able to simulate bremsstrahlung emission by electrons with energies E not included in the grid.

Sampling of the photon energy

The PDF for the reduced photon energy, $\kappa = W/E$, is given by [see eq. (3.126)]

$$p(E,\kappa) = \frac{1}{\kappa}\chi(Z,E,\kappa)\,\Theta(\kappa-\kappa_{\rm cr}), \qquad (3.147)$$

where $\kappa_{\rm cr} = W_{\rm cr}/E$ and $\chi(Z,E,\kappa)$ is calculated by linear interpolation, in both $\ln E$ and κ, in the stored table. That is, $\chi(Z,E,\kappa)$ is considered to be a piecewise linear function

of κ. To sample κ from the PDF (3.147) for an energy E_i in the grid, we express the interpolated scaled DCS as

$$\chi(Z, E_i, \kappa) = a_j + b_j \kappa \qquad \text{if } \kappa_j \leq \kappa \leq \kappa_{j+1}, \tag{3.148}$$

and introduce the cumulative distribution function,

$$\mathcal{P}_j = \int_{\kappa_{cr}}^{\kappa_j} p(E_i, \kappa)\, \mathrm{d}\kappa, \tag{3.149}$$

which, for a piecewise linear χ, can be computed exactly. We also define

$$\chi_{\max,j} = \max\left\{\chi(Z, E, \kappa), \kappa \in (\kappa_j, \kappa_{j+1})\right\} \qquad j = 1, \ldots, 32. \tag{3.150}$$

With all this we can formulate the following sampling algorithm, which combines a numerical inverse transform and a rejection,

(i) Generate a random number ξ and determine the index j for which $\mathcal{P}_j \leq \xi \mathcal{P}_{32} \leq \mathcal{P}_{j+1}$ using the binary search method.

(ii) Sample κ from the distribution κ^{-1} in the interval (κ_j, κ_{j+1}), i.e.

$$\kappa = \kappa_j \left(\kappa_{j+1}/\kappa_j\right)^{\xi}. \tag{3.151}$$

(iii) If $\xi \chi_{\max,j} < a_j + b_j \kappa$, deliver κ.

(iv) Go to step (i).

This sampling algorithm is exact and very fast [notice that the binary search in step (i) requires at most 5 comparisons], but is only applicable for the energies in the grid where χ is tabulated.

To simulate bremsstrahlung emission by electrons with energies E not included in the grid, we should first obtain the PDF $p(E, \kappa)$ by interpolation along the energy axis and then perform the random sampling of κ from this PDF using the algorithm described above. This procedure is too time consuming. A faster method consists of assuming that the grid of energies is dense enough so that linear interpolation in $\ln E$ is sufficiently accurate. If $E_i < E < E_{i+1}$, we can express the interpolated PDF as

$$p_{\text{int}}(E, \kappa) = \pi_i\, p(E_i, \kappa) + \pi_{i+1}\, p(E_{i+1}, \kappa) \tag{3.152}$$

with

$$\pi_i = \frac{\ln E_{i+1} - \ln E}{\ln E_{i+1} - \ln E_i}, \qquad \pi_{i+1} = \frac{\ln E - \ln E_i}{\ln E_{i+1} - \ln E_i}. \tag{3.153}$$

These "interpolation weights" are positive and add to unity, i.e. they can be interpreted as point probabilities. Therefore, to perform the random sampling of κ from $p_{\text{int}}(E, \kappa)$ we can employ the composition method (section 1.2.5), which leads to the following algorithm:

(i) Sample the integer variable I, which can take the values i or $i+1$ with point probabilities π_i and π_{i+1}, respectively.

(ii) Sample κ from the distribution $p_{\text{int}}(E_I;\kappa)$.

With this "interpolation by weight" method we only need to sample κ from the tabulated PDFs, i.e. for the energies E_i of the grid.

Angular distribution of emitted photons

The random sampling of $\cos\theta$ is simplified by noting that the PDF given by eq. (3.146) results from a Lorentz transformation, with speed β', of the PDF (3.143). This means that we can sample the photon direction $\cos\theta'$ in the reference frame K′ from the PDF (3.143) and then apply the transformation (3.144) (with β' instead of β) to get the direction $\cos\theta$ in the laboratory frame.

To generate random values of $\cos\theta$ from (3.146) we use the following algorithm, which combines the composition and rejection methods,

(i) Sample a random number ξ_1.

(ii) If $\xi_1 < A$, then
1) Sample a random number ξ and set $\cos\theta' = -1 + 2\xi$.
2) Sample a random number ξ.
3) If $2\xi > 1 + \cos^2\theta'$, go to 1).

(iii) If $\xi_1 \geq A$, then
4) Sample a random number ξ and set $\cos\theta' = -1 + 2\xi$.
5) Sample a random number ξ.
6) If $\xi > 1 - \cos^2\theta'$, go to 4).

(iv) Deliver $\cos\theta = \dfrac{\cos\theta' + \beta'}{1 + \beta'\cos\theta'}$.

The efficiencies of the rejections in steps (ii) and (iii) are both equal to 0.66. That is, on average, we need 4 random numbers to generate each value of $\cos\theta$.

3.4 Positron annihilation

Following Nelson et al. (1985), we consider that positrons penetrating a medium of atomic number Z with kinetic energy E can annihilate with the electrons in the medium by emission of two photons. We assume that the target electrons are free and at rest, thus disregarding electron binding effects, which enable one-photon annihilation (Heitler, 1954). When annihilation occurs in flight, i.e. when the kinetic energy E of the positron is larger than the "absorption" energy, the two photons may have different energies, say

E_- and E_+, which add to $E+2\mathrm{m_e}c^2$. In what follows, quantities referring to the photon with the lowest energy will be denoted by the subscript "$-$". Each annihilation event is then completely characterized by the quantity

$$\zeta \equiv \frac{E_-}{E + 2\mathrm{m_e}c^2}. \tag{3.154}$$

Assuming that the positron moves initially in the direction of the z-axis, from conservation of energy and momentum it follows that the two photons are emitted in directions with polar angles [see eqs. (A.21) and (A.22) in appendix A]

$$\cos\theta_- = (\gamma^2 - 1)^{-1/2}(\gamma + 1 - 1/\zeta) \tag{3.155}$$

and

$$\cos\theta_+ = (\gamma^2 - 1)^{-1/2}[\gamma + 1 - 1/(1-\zeta)], \tag{3.156}$$

and azimuthal angles ϕ_- and $\phi_+ = \phi_- + \pi$. The quantity $\gamma = 1 + E/(\mathrm{m_e}c^2)$ is the total energy of the positron in units of its rest energy.

The maximum value of ζ is $1/2$, its minimum value is found when $\cos\theta_- = -1$ and is given by

$$\zeta_{\mathrm{min}} = \frac{1}{\gamma + 1 + (\gamma^2 - 1)^{1/2}}. \tag{3.157}$$

The DCS (per electron) for two-photon annihilation, as observed in the centre-of-mass system of the positron and the electron, is given by Heitler (1954). Nelson et al. (1985) transformed this DCS to the laboratory system (where the electron is at rest), their result can be written as

$$\frac{\mathrm{d}\sigma_{\mathrm{an}}}{\mathrm{d}\zeta} = \frac{\pi r_\mathrm{e}^2}{(\gamma+1)(\gamma^2-1)}\left[S(\zeta) + S(1-\zeta)\right], \tag{3.158}$$

where

$$S(\zeta) = -(\gamma+1)^2 + (\gamma^2 + 4\gamma + 1)\frac{1}{\zeta} - \frac{1}{\zeta^2}. \tag{3.159}$$

Owing to the axial symmetry of the process, the DCS is independent of the azimuthal angle ϕ_-, which is uniformly distributed on the interval $(0, 2\pi)$. For fast positrons, annihilation photons are emitted preferentially at forward directions. When the kinetic energy of the positron decreases, the angular distribution of the generated photons becomes more isotropical (see fig. 3.16).

The cross section (per target electron) for two-photon annihilation is

$$\sigma_{\mathrm{an}} = \int_{\zeta_{\mathrm{min}}}^{1/2} \frac{\mathrm{d}\sigma_{\mathrm{an}}}{\mathrm{d}\zeta}\,\mathrm{d}\zeta = \frac{\pi r_\mathrm{e}^2}{(\gamma+1)(\gamma^2-1)} \times \left\{(\gamma^2 + 4\gamma + 1)\ln\left[\gamma + \left(\gamma^2-1\right)^{1/2}\right] - (3+\gamma)\left(\gamma^2-1\right)^{1/2}\right\}. \tag{3.160}$$

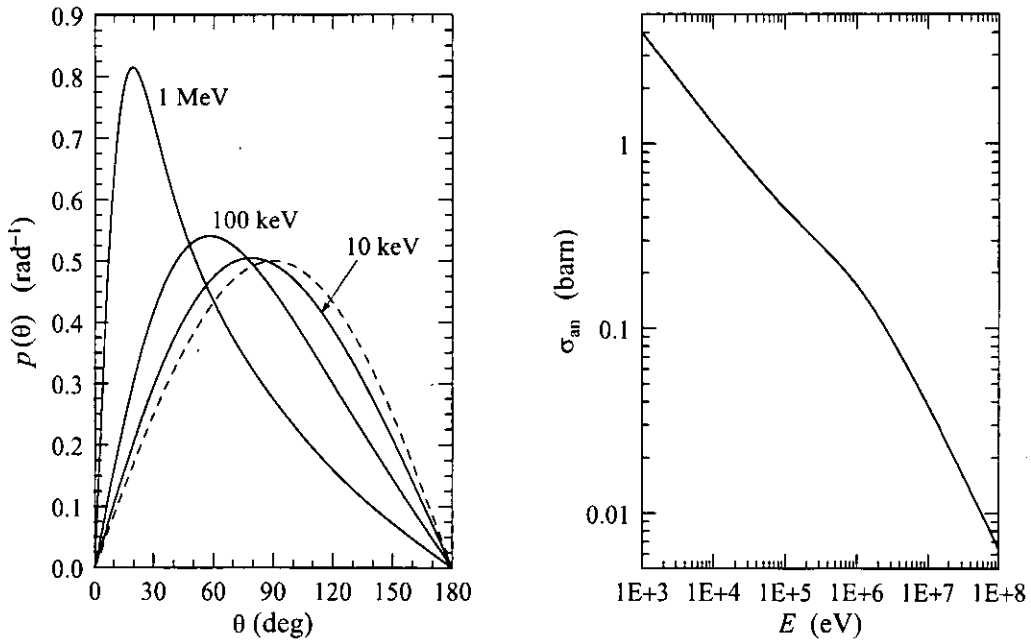


Figure 3.16: Left: angular distributions of photons produced by annihilation in flight of positrons with the indicated kinetic energies. The dashed line represents the isotropic distribution. Right: Annihilation cross section per target electron as a function of the kinetic energy of the positron.

The annihilation mean free path is given by

$$\lambda_{\rm an}^{-1} = \mathcal{N}Z\sigma_{\rm an}, \tag{3.161}$$

where $\mathcal{N}Z$ is the density of electrons in the medium. The annihilation cross section is displayed in fig. 3.16. The cross section decreases with the kinetic energy and, therefore, high-energy positrons can travel path lengths of the order of the CSDA range before annihilating.

3.4.1 Generation of emitted photons

The PDF of ζ is given by (normalization is irrelevant here)

$$p_{\rm an}(\zeta) = S(\zeta) + S(1-\zeta), \qquad \zeta_{\rm min} \le \zeta \le 1/2. \tag{3.162}$$

To sample ζ, we may take advantage of the symmetry of this expression under the exchange of the two photons, which corresponds to exchanging ζ and $1-\zeta$. We first consider the distribution

$$P(v) \equiv S(v), \qquad \zeta_{\rm min} \le v \le 1-\zeta_{\rm min} \tag{3.163}$$

and write it in the form

$$P(\upsilon) = \pi(\upsilon) g(\upsilon) \tag{3.164}$$

with

$$\pi(\upsilon) = \left[\ln\left(\frac{1-\zeta_{\min}}{\zeta_{\min}}\right)\right]^{-1} \frac{1}{\upsilon} \tag{3.165}$$

and

$$g(\upsilon) = \left[-(\gamma+1)^2\upsilon + (\gamma^2+4\gamma+1) - \frac{1}{\upsilon}\right]. \tag{3.166}$$

$\pi(\upsilon)$ is a proper PDF (i.e. it is definite positive and normalized to unity) and $g(\upsilon)$ is a monotonically decreasing function. Random values of υ from the distribution $P(\upsilon)$ can be generated by using the following algorithm (rejection method):

(i) Sample a value υ from the distribution $\pi(\upsilon)$. This is easily done with the inverse transform method, which yields the following sampling equation

$$\upsilon = \zeta_{\min}\left(\frac{1-\zeta_{\min}}{\zeta_{\min}}\right)^{\xi}. \tag{3.167}$$

(ii) Generate a new random number ξ.

(iii) If $\xi g(\zeta_{\min}) > g(\upsilon)$, go to step (i).

(iv) Deliver υ.

It is clear that the random value

$$\zeta = \min(\upsilon, 1-\upsilon) \tag{3.168}$$

follows the distribution given by eq. (3.162) when υ is sampled from the distribution $P(\upsilon)$. The efficiency of this sampling algorithm practically equals 100% for positrons with kinetic energy E less than 10 keV, decreases when E increases to reach a minimum value of $\sim 80\%$ at $E \sim 10$ MeV and increases monotonically for larger energies.

As the result of annihilation, two photons with energies $E_- = \zeta(E + 2m_e c^2)$ and $E_+ = (1-\zeta)(E + 2m_e c^2)$ are emitted in the directions given by eqs. (3.155) and (3.156).

Chapter 4

Electron/positron transport mechanics

In principle, the scattering model and sampling techniques described in chapter 3 allows the detailed Monte Carlo simulation of electron and positron transport in matter. However, detailed simulation is feasible only when the mean number of interactions per track is small (a few hundred at most). This occurs for electrons with low initial kinetic energies or for thin geometries. The number of interactions experienced by an electron or positron before being effectively stopped increases with its initial energy and, therefore, detailed simulation becomes impractical at high energies.

PENELOPE implements a "mixed" simulation scheme (Berger, 1963; Reimer and Krefting, 1976; Andreo and Brahme, 1984), which combines the detailed simulation of hard events (i.e. events with polar angular deflection θ or energy loss W larger than previously selected cutoff values θ_c and W_c) with condensed simulation of soft events, in which $\theta < \theta_c$ or $W < W_c$. Owing to the fact that for high-energy electrons the DCSs for the various interaction processes decrease rapidly with the polar scattering angle and the energy loss, cutoff values can be selected such that the mean number of hard events per electron track is sufficiently small to permit their detailed simulation. In general, this is accomplished by using relatively small cutoff values, so that each soft interaction has only a slight effect on the simulated track. The global effect of the (usually many) soft interactions that take place between each pair of consecutive hard events can then be simulated accurately by using a multiple scattering approach. Hard events occur much less frequently than soft events, but they have severe effects on the track evolution (i.e. they cause large angular deflections and lateral displacements or considerable energy losses), which can only be properly reproduced by detailed simulation. The computer time needed to simulate each track diminishes rapidly when the cutoff values for the angular deflection and the energy loss are increased. Mixed simulation algorithms are usually very stable under variations of the adopted cutoff values, whenever these are kept below some reasonable limits. Mixed simulation is then preferable to condensed simulation because 1) spatial distributions are simulated more accurately, 2) tracks in the vicinity of interfaces are properly handled, and 3) possible dependence of the results on user-defined parameters is largely reduced.

4.1 Elastic scattering

Let us start by considering electrons (or positrons) with kinetic energy E moving in a hypothetical infinite homogeneous medium, with $\mathcal{N}$ scattering centres per unit volume, in which they experience only pure elastic collisions (i.e. with no energy loss).

4.1.1 Multiple elastic scattering theory

Assume that an electron starts off from a certain position, which we select as the origin of our reference frame, moving in the direction of the z-axis. Let $f(s;\mathbf{r},\hat{\mathbf{d}})$ denote the probability density of finding the electron at the position $\mathbf{r}=(x,y,z)$, moving in the direction given by the unit vector $\hat{\mathbf{d}}$ after having travelled a path length s. The diffusion equation for this problem is (Lewis, 1950)

$$\frac{\partial f}{\partial s}+\hat{\mathbf{d}}\cdot\nabla f=\mathcal{N}\int\left[f(s;\mathbf{r},\hat{\mathbf{d}}')-f(s;\mathbf{r},\hat{\mathbf{d}})\right]\frac{\mathrm{d}\sigma_{\mathrm{el}}(\theta)}{\mathrm{d}\Omega}\,\mathrm{d}\Omega, \tag{4.1}$$

where $\theta\equiv\arccos(\hat{\mathbf{d}}\cdot\hat{\mathbf{d}}')$ is the scattering angle corresponding to the angular deflection $\hat{\mathbf{d}}'\rightarrow\hat{\mathbf{d}}$. This equation has to be solved with the boundary condition $f(0;\mathbf{r},\hat{\mathbf{d}})=(1/\pi)\delta(\mathbf{r})\delta(1-\cos\chi)$, where χ is the polar angle of the direction $\hat{\mathbf{d}}$. By expanding $f(s;\mathbf{r},\hat{\mathbf{d}})$ in spherical harmonics, Lewis (1950) obtained exact expressions for the angular distribution and for the first moments of the spatial distribution after a given path length s. The probability density $F(s;\chi)$ of having a final direction in the solid angle element $\mathrm{d}\Omega$ around a direction defined by the polar angle χ is given by

$$F(s;\chi)=\int f(s;\mathbf{r},\hat{\mathbf{d}})\,\mathrm{d}\mathbf{r}=\sum_{\ell=0}^{\infty}\frac{2\ell+1}{4\pi}\exp(-s/\lambda_{\mathrm{el},\ell})P_\ell(\cos\chi), \tag{4.2}$$

where $P_\ell(\cos\chi)$ are Legendre polynomials and $\lambda_{\mathrm{el},\ell}=1/(\mathcal{N}\sigma_{\mathrm{el},\ell})$ is the ℓ-th transport mean free path defined by eq. (3.14). The result given by eq. (4.2) coincides with the multiple scattering distribution obtained by Goudsmit and Saunderson (1940a, 1940b). Evidently, the distribution $F(s;\chi)$ is symmetric about the z-axis, i.e. independent of the azimuthal angle of the final direction.

From the orthogonality of the Legendre polynomials, it follows that

$$\langle P_\ell(\cos\chi)\rangle\equiv 2\pi\int_{-1}^{1}P_\ell(\cos\chi)F(s;\chi)\,\mathrm{d}(\cos\chi)=\exp(-s/\lambda_{\mathrm{el},\ell}). \tag{4.3}$$

In particular, we have

$$\langle\cos\chi\rangle=\exp(-s/\lambda_{\mathrm{el},1}) \tag{4.4}$$

and

$$\langle\cos^2\chi\rangle=\frac{1}{3}\left[1+2\exp(-s/\lambda_{\mathrm{el},2})\right]. \tag{4.5}$$

Lewis (1950) also derived analytical formulae for the first moments of the spatial distribution and the correlation function of z and $\cos\chi$. Neglecting energy losses, the results explicitly given in Lewis' paper simplify to

$$\langle z \rangle \equiv 2\pi \int z f(s;\mathbf{r},\hat{\mathbf{d}})\,\mathrm{d}(\cos\chi)\,\mathrm{d}\mathbf{r} = \lambda_{\mathrm{el},1}\left[1-\exp(-s/\lambda_{\mathrm{el},1})\right], \tag{4.6}$$

$$\begin{aligned}\langle x^2+y^2 \rangle &\equiv 2\pi \int \left(x^2+y^2\right) f(s;\mathbf{r},\hat{\mathbf{d}})\,\mathrm{d}(\cos\chi)\,\mathrm{d}\mathbf{r} \\ &= \frac{4}{3}\int_0^s \mathrm{d}t\,\exp(-t/\lambda_{\mathrm{el},1})\int_0^t \left[1-\exp(-u/\lambda_{\mathrm{el},2})\right]\exp(u/\lambda_{\mathrm{el},1})\,\mathrm{d}u,\end{aligned} \tag{4.7}$$

$$\begin{aligned}\langle z\cos\chi \rangle &\equiv 2\pi \int z\cos\chi f(s;\mathbf{r},\hat{\mathbf{d}})\,\mathrm{d}(\cos\chi)\,\mathrm{d}\mathbf{r} \\ &= \exp(-s/\lambda_{\mathrm{el},1})\int_0^s \left[1+2\exp(-t/\lambda_{\mathrm{el},2})\right]\exp(t/\lambda_{\mathrm{el},1})\,\mathrm{d}t.\end{aligned} \tag{4.8}$$

It is worth observing that the quantities (4.4)–(4.8) are completely determined by the values of the transport mean free paths $\lambda_{\mathrm{el},1}$ and $\lambda_{\mathrm{el},2}$; they are independent of the elastic mean free path λ_{el}.

4.1.2 Mixed simulation of elastic scattering

At high energies, where detailed simulation becomes impractical, $\lambda_{\mathrm{el},1} \gg \lambda_{\mathrm{el}}$ (see fig. 3.3) so that the average angular deflection in each collision is small. In other words, the great majority of elastic collisions of fast electrons are soft collisions with very small deflections. We shall consider mixed simulation procedures (see Fernández-Varea et al., 1993b; Baró et al., 1994b) in which hard collisions, with scattering angle θ larger than a certain value θ_{c}, are individually simulated and soft collisions (with $\theta < \theta_{\mathrm{c}}$) are described by means of a multiple scattering approach.

In practice, the mixed algorithm will be defined by specifying the mean free path $\lambda_{\mathrm{el}}^{(\mathrm{h})}$ between hard elastic events, defined by [see eqs. (3.10) and (3.12)]

$$\frac{1}{\lambda_{\mathrm{el}}^{(\mathrm{h})}} = \mathcal{N}2\pi\int_{\theta_{\mathrm{c}}}^{\pi}\frac{\mathrm{d}\sigma_{\mathrm{el}}(\theta)}{\mathrm{d}\Omega}\sin\theta\,\mathrm{d}\theta. \tag{4.9}$$

This equation determines the cutoff angle θ_{c} as a function of $\lambda_{\mathrm{el}}^{(\mathrm{h})}$. A convenient recipe to set the mean free path $\lambda_{\mathrm{el}}^{(\mathrm{h})}$ is

$$\lambda_{\mathrm{el}}^{(\mathrm{h})}(E) = \max\left\{\lambda_{\mathrm{el}}(E), C_1\lambda_{\mathrm{el},1}(E)\right\}, \tag{4.10}$$

where C_1 is a pre-selected small constant (say, less than ~ 0.1). For increasing energies, λ_{el} attains a constant value and $\lambda_{\mathrm{el},1}$ increases steadily (see fig. 3.3) so that the formula (4.10) gives a mean free path for hard collisions that increases with energy, i.e. hard collisions are less frequent when the scattering effect is weaker. The recipe (4.10) also ensures that $\lambda_{\mathrm{el}}^{(\mathrm{h})}$ will reduce to the actual mean free path λ_{el} for low energies. In this case,

soft collisions cease to occur ($\theta_c = 0$) and mixed simulation becomes purely detailed. It is worth noticing that, when mixed simulation is effective (i.e. when $\lambda_{el}^{(h)} > \lambda_{el}$), the mean angular deflection in a path length $\lambda_{el}^{(h)}$ is [see eq. (4.4)]

$$1 - \langle \cos\chi \rangle = 1 - \exp(-\lambda_{el}^{(h)}/\lambda_{el,1}) \simeq C_1. \tag{4.11}$$

Hence, when using the prescription (4.10), the average angular deflection due to all elastic collisions occurring along a path length $\lambda_{el}^{(h)}$ equals C_1.

The PDF of the step length s between two successive hard collisions is

$$p(s) = \frac{1}{\lambda_{el}^{(h)}} \exp(-s/\lambda_{el}^{(h)}), \tag{4.12}$$

and random values of s can be generated by means of the sampling formula, eq. (1.36)

$$s = -\lambda_{el}^{(h)} \ln\xi. \tag{4.13}$$

The (unnormalized) PDF of the polar deflection θ in single hard collisions is

$$p^{(h)}(\theta) = \frac{d\sigma_{el}(\theta)}{d\Omega} \sin\theta\, \Theta(\theta - \theta_c), \tag{4.14}$$

where $\Theta(x)$ stands for the step function.

The inverse transport mean free paths $\lambda_{el,\ell}^{-1}$, see eq. (3.14), for the actual scattering process can be split into contributions from soft and hard collisions, i.e.

$$\frac{1}{\lambda_{el,\ell}} = \frac{1}{\lambda_{el,\ell}^{(s)}} + \frac{1}{\lambda_{el,\ell}^{(h)}}, \tag{4.15}$$

where

$$\frac{1}{\lambda_{el,\ell}^{(s)}} = \mathcal{N} 2\pi \int_0^{\theta_c} [1 - P_\ell(\cos\theta)] \frac{d\sigma_{el}(\theta)}{d\Omega} \sin\theta\, d\theta \tag{4.16a}$$

and

$$\frac{1}{\lambda_{el,\ell}^{(h)}} = \mathcal{N} 2\pi \int_{\theta_c}^{\pi} [1 - P_\ell(\cos\theta)] \frac{d\sigma_{el}(\theta)}{d\Omega} \sin\theta\, d\theta. \tag{4.16b}$$

Let us assume that an electron starts off from the origin of coordinates moving in the direction of the z-axis and undergoes the first hard collision after travelling a path length s. The exact angular distribution produced by the soft collisions along this step is

$$F^{(s)}(s;\chi) = \sum_{\ell=0}^{\infty} \frac{2\ell+1}{4\pi} \exp(-s/\lambda_{el,\ell}^{(s)}) P_\ell(\cos\chi). \tag{4.17}$$

The exact average longitudinal and transverse displacements at the end of the step are given by [see eqs. (4.6) and (4.7)]

$$\langle z \rangle^{(s)} = \lambda_{el,1}^{(s)} \left[1 - \exp(-s/\lambda_{el,1}^{(s)})\right] = s \left[1 - \frac{1}{2}\left(\frac{s}{\lambda_{el,1}^{(s)}}\right) + \frac{1}{6}\left(\frac{s}{\lambda_{el,1}^{(s)}}\right)^2 - \ldots\right], \tag{4.18}$$

$$\langle x^2+y^2\rangle^{(s)} = \frac{2}{9}\frac{s^3}{\lambda_{el,2}^{(s)}}\left[1-\frac{1}{4}\left(1+\frac{\lambda_{el,1}^{(s)}}{\lambda_{el,2}^{(s)}}\right)\left(\frac{s}{\lambda_{el,1}^{(s)}}\right)+\ldots\right], \tag{4.19}$$

where $\lambda_{el,1}^{(s)}$, the first transport mean free path for soft collisions, is larger than $\lambda_{el,1}$. As the mean free path between hard collisions is normally much less than $\lambda_{el,1}^{(s)}$ (depending on the value of C_1), the value $s/\lambda_{el,1}^{(s)}$ is, on average, much less than unity (note that $\langle s\rangle = \lambda_{el}^{(h)}$). Therefore, the global effect of the soft collisions in the step, i.e. the change in direction of movement *and* the lateral displacement, is very small (part of the deflection is caused by the hard interaction at the end of the step).

In PENELOPE, the angular deflection and the lateral displacement due to the multiple soft collisions in a step of length s are simulated by means of the random hinge method[1] (Fernández-Varea et al., 1993b). The associated algorithm can be formulated as follows (see fig. 4.1),

(i) The electron first moves a random distance τ, which is sampled uniformly in the interval $(0, s)$, in the initial direction.

(ii) Then a single artificial soft scattering event (a hinge) takes place, in which the electron changes its direction of movement according to the multiple scattering distribution $F^{(s)}(s;\chi)$.

(iii) Finally, the electron moves a distance $s-\tau$ in the new direction.

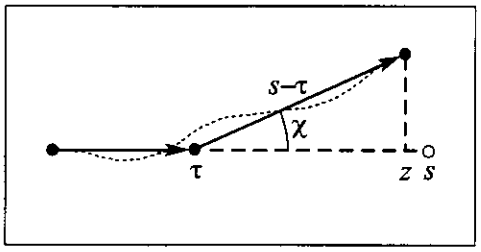


Figure 4.1: Simulation of the global effect of soft collisions between two consecutive hard collisions by the random hinge method.

Obviously, this algorithm leads to the exact angular distribution at the end of the step. The average longitudinal displacement at the end of the simulated step is

$$\langle z\rangle_{sim}^{(s)} = \frac{s}{2}+\frac{s}{2}\langle\cos\chi\rangle^{(s)} = s\left[1-\frac{1}{2}\left(\frac{s}{\lambda_{el,1}^{(s)}}\right)+\frac{1}{4}\left(\frac{s}{\lambda_{el,1}^{(s)}}\right)^2-\ldots\right], \tag{4.20}$$

which agrees closely with the exact result given by eq. (4.18). Moreover, the average simulated transverse displacement is

$$\langle x^2+y^2\rangle_{sim}^{(s)} = \langle(s-\tau)^2\sin^2\chi\rangle_{sim}^{(s)} = \frac{1}{3}s^2\left(1-\langle\cos^2\chi\rangle^{(s)}\right)$$

[1]The name was coined by Ron Kensek.

$$= \frac{2}{9}\frac{s^3}{\lambda_{\mathrm{el},2}^{(s)}}\left[1 - \frac{1}{2}\frac{\lambda_{\mathrm{el},1}^{(s)}}{\lambda_{\mathrm{el},2}^{(s)}}\left(\frac{s}{\lambda_{\mathrm{el},1}^{(s)}}\right) + \ldots\right], \tag{4.21}$$

which does not differ much from the exact value given by eq. (4.19). From these facts, we may conclude that the random hinge method provides a faithful description of the transport when the step length s is much shorter than the first transport mean free path $\lambda_{\mathrm{el},1}$, so that the global angular deflection and lateral displacement are small. Surprisingly, it does work well also in condensed (class I) simulations, where this requirement is not met. In spite of its simplicity, the random hinge method competes in accuracy and speed with other, much more sophisticated transport algorithms (see Bielajew and Salvat, 2001, and references therein). It seems that the randomness of the hinge position τ leads to correlations between the angular deflection and the displacement that are close to the actual correlations.

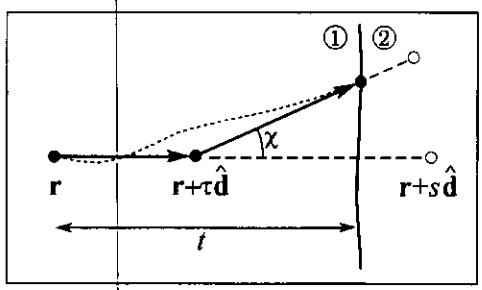


Figure 4.2: Simulation of a track near the crossing of an interface.

The random hinge algorithm can be readily adapted to simulate multiple elastic scattering processes in limited material structures, which may consist of several regions of different compositions separated by well-defined surfaces (interfaces). In these geometries, when the track crosses an interface, we simply stop it at the crossing point, and resume the simulation in the new material. In spite of its simplicity, this recipe gives a fairly accurate description of interface crossing. To see this, consider that a hard collision has occurred at the position $\mathbf{r}$ in region "1" and assume that the following hard collision occurs in region "2". The step length s between these two hard collisions is larger than the distance t from $\mathbf{r}$ to the interface (see fig. 4.2). If the artificial soft elastic collision occurs in region "1", the angular deflection in this collision is sampled from the distribution $F^{(s)}(s;\chi)$. Otherwise, the electron reaches the interface without changing its direction of movement. Assuming $s \ll \lambda_{\mathrm{el},1}^{(s)}$, the mean angular deflection in a soft collision is

$$1 - \langle\cos\chi\rangle^{(s)} = 1 - \exp(-s/\lambda_{\mathrm{el},1}^{(s)}) \simeq \frac{s}{\lambda_{\mathrm{el},1}^{(s)}}. \tag{4.22}$$

Moreover, when this assumption is valid, lateral displacements due to soft collisions are small and can be neglected to a first approximation. As the probability for the soft collision to occur within region "1" equals t/s, the average angular deflection of the

simulated electron track when it reaches the interface is

$$1 - \langle \cos\chi \rangle = \frac{t}{s}\left(1 - \langle \cos\chi \rangle^{(s)}\right) \simeq \frac{t}{\lambda_{\mathrm{el},1}^{(s)}}, \tag{4.23}$$

which practically coincides with the exact mean deviation after the path length t within region "1", as required. Thus, by sampling the position of the soft collision uniformly in the segment $(0, s)$ we make sure that the electron reaches the interface with the correct average direction of movement.

Angular deflections in soft scattering events

In the random hinge method, the global effect of the soft collisions experienced by the particle along a path segment of length s between two consecutive hard events is simulated as a single artificial soft scattering event. The angular deflection follows the multiple scattering distribution $F^{(s)}(s;\chi)$. Unfortunately, the exact Legendre expansion, eq. (4.17), is not appropriate for Monte Carlo simulation, since this expansion converges very slowly (because the associated single scattering DCS is not continuous) and the sum varies rapidly with the path length s.

Whenever the cutoff angle θ_c is small, the distribution $F^{(s)}(s;\chi)$ may be calculated by using the small angle approximation (see e.g. Lewis, 1950). Notice that θ_c can be made as small as desired by selecting a small enough value of C_1, see eqs. (4.9) and (4.10). Introducing the limiting form of the Legendre polynomials

$$P_\ell(\cos\theta) \simeq 1 - \frac{1}{4}\ell(\ell+1)\theta^2 \tag{4.24}$$

into eq. (4.16a) we get

$$\frac{1}{\lambda_{\mathrm{el},\ell}^{(s)}} = \mathcal{N} 2\pi \frac{\ell(\ell+1)}{4}\int_0^{\theta_c} \theta^2 \frac{\mathrm{d}\sigma_{\mathrm{el}}(\theta)}{\mathrm{d}\Omega}\sin\theta\,\mathrm{d}\theta = \frac{\ell(\ell+1)}{2}\frac{1}{\lambda_{\mathrm{el},1}^{(s)}}, \tag{4.25}$$

i.e. the transport mean free paths $\lambda_{\mathrm{el},\ell}^{(s)}$ are completely determined by the single value $\lambda_{\mathrm{el},1}^{(s)}$. The angular distribution $F^{(s)}$ then simplifies to

$$F^{(s)}(s;\chi) = \sum_{\ell=0}^{\infty} \frac{2\ell+1}{4\pi}\exp\left[-\frac{\ell(\ell+1)}{2}\frac{s}{\lambda_{\mathrm{el},1}^{(s)}}\right] P_\ell(\cos\chi). \tag{4.26}$$

This expression can be evaluated by using the Molière (1948) approximation for the Legendre polynomials, we obtain (see Fernández-Varea et al., 1993b)

$$F^{(s)}(s;\chi) = \frac{1}{2\pi}\left(\frac{\chi}{\sin\chi}\right)^{1/2}\frac{\lambda_{\mathrm{el},1}^{(s)}}{s}\exp\left[\frac{s}{8\lambda_{\mathrm{el},1}^{(s)}} - \frac{\lambda_{\mathrm{el},1}^{(s)}}{2s}\chi^2\right], \tag{4.27}$$

which does not differ significantly from the Gaussian distribution with variance $s/\lambda_{\mathrm{el},1}^{(s)}$. This result is accurate whenever $s \ll \lambda_{\mathrm{el},1}^{(s)}$ and $\theta_c \ll 1$. It offers a possible method

of generating the angular deflection in artificial soft events. When the result given by eq. (4.27) is applicable, the single parameter $\lambda_{\mathrm{el},1}^{(\mathrm{s})}$ completely determines the multiple scattering distribution due to soft collisions, i.e. other details of the DCS for scattering angles less than θ_{c} are irrelevant. However, in actual Monte Carlo simulations, the small-angle approximation is seldom applicable.

In most practical cases the number of hard collisions per electron track can be made relatively large by simply using a small value of the parameter C_1 [see eq. (4.10)]. When the number of steps is large enough, say larger than ~ 10, it is not necessary to use the exact distribution $F^{(\mathrm{s})}(s;\chi)$ to sample the angular deflection in artificial soft collisions. Instead, we may use a simpler distribution, $F_{\mathrm{a}}(s;\chi)$, with the same mean and variance, without appreciably distorting the simulation results. This is so because the details of the adopted distribution are washed out after a sufficiently large number of steps and will not be seen in the simulated distributions. Notice that, within the small angle approximation, it is necessary to keep only the proper value of the first moment to get the correct final distributions. However, if the cutoff angle θ_{c} is not small enough, the angular distribution $F^{(\mathrm{s})}(s;\chi)$ may become sensitive to higher-order moments of the soft single scattering distribution. Thus, by also keeping the proper value of the variance, the range of validity of the simulation algorithm is extended, i.e. we can speed up the simulation by using larger values of C_1 (or of $\lambda_{\mathrm{el}}^{(\mathrm{h})}$) and still obtain the correct distributions.

We now return to the notation of section 3.1, and use the variable $\mu \equiv (1-\cos\chi)/2$ to describe angular deflections in soft scattering events. The exact first and second moments of the multiple scattering distribution $F^{(\mathrm{s})}(s;\mu)$ are

$$\langle\mu\rangle^{(\mathrm{s})} \equiv \int_0^1 \mu F_{\mathrm{a}}(s;\mu)\,\mathrm{d}\mu = \frac{1}{2}\left[1-\exp(-s/\lambda_{\mathrm{el},1}^{(\mathrm{s})})\right] \tag{4.28}$$

and

$$\langle\mu^2\rangle^{(\mathrm{s})} \equiv \int_0^1 \mu^2 F_{\mathrm{a}}(s;\mu)\,\mathrm{d}\mu = \langle\mu\rangle^{(\mathrm{s})} - \frac{1}{6}\left[1-\exp(-s/\lambda_{\mathrm{el},2}^{(\mathrm{s})})\right]. \tag{4.29}$$

The angular deflection in soft scattering events will be generated from a distribution $F_{\mathrm{a}}(s;\mu)$, which is required to satisfy eqs. (4.28) and (4.29), but is otherwise arbitrary. PENELOPE uses the following,

$$F_{\mathrm{a}}(s;\mu) = aU_{0,b}(\mu) + (1-a)U_{b,1}(\mu), \tag{4.30}$$

where $U_{u,v}(x)$ denotes the normalized uniform distribution in the interval (u,v),

$$U_{u,v}(x) = \begin{cases} 1/(v-u) & \text{if } u \le x \le v, \\ 0 & \text{otherwise.} \end{cases} \tag{4.31}$$

The parameters a and b, obtained from the conditions (4.28) and (4.29), are

$$b = \frac{2\langle\mu\rangle^{(\mathrm{s})} - 3\langle\mu^2\rangle^{(\mathrm{s})}}{1-2\langle\mu\rangle^{(\mathrm{s})}}, \qquad a = 1-2\langle\mu\rangle^{(\mathrm{s})}+b. \tag{4.32}$$

The simple distribution (4.30) is flexible enough to reproduce the combinations of first and second moments encountered in the simulations [notice that $\langle\mu\rangle^{(s)}$, eq. (4.28), is always less than 1/2] and allows fast random sampling of μ.

4.1.3 Simulating with the MW model

PENELOPE simulates elastic scattering by using the MW model (see section 3.1), which allows the formulation of the mixed simulation algorithm in closed analytical form.

The mean free path $\lambda_{\rm el}^{(h)}$ between hard elastic events and the cutoff deflection $\mu_{\rm c} = (1-\cos\theta_{\rm c})/2$ are related through [see eqs. (3.18) and (4.9)]

$$\frac{1}{\lambda_{\rm el}^{(h)}} = \frac{1}{\lambda_{\rm el}} \int_{\mu_{\rm c}}^{1} p_{\rm MW}(\mu)\, {\rm d}\mu. \tag{4.33}$$

This equation can be easily inverted to give

$$\mu_{\rm c} = \mathcal{P}_{\rm MW}^{-1}(\xi_{\rm c}), \tag{4.34}$$

where

$$\xi_{\rm c} \equiv 1 - \frac{\lambda_{\rm el}}{\lambda_{\rm el}^{(h)}} \tag{4.35}$$

and $\mathcal{P}_{\rm MW}^{-1}$ is the inverse of the single scattering cumulative distribution function given by eqs. (3.31) and (3.36).

In the following, we assume that the MW distribution is that of case I, eq. (3.24); the formulae for case II can be derived in a similar way. The random sampling of the angular deflection μ in hard collisions is performed by the inverse transform method (section 1.2.2); random values of μ are obtained from the sampling equation

$$\int_{\mu_{\rm c}}^{\mu} p_{\rm MW}(\mu')\, {\rm d}\mu' = \xi \int_{\mu_{\rm c}}^{1} p_{\rm MW}(\mu')\, {\rm d}\mu'. \tag{4.36}$$

With the MW distribution, eq. (3.24), this equation can be solved analytically to give

$$\mu = \mathcal{P}_{\rm MW}^{-1}\left(1 - \frac{\lambda_{\rm el}}{\lambda_{\rm el}^{(h)}}(1-\xi)\right). \tag{4.37}$$

To determine the angular distribution of soft events $F_{\rm a}(s;\mu)$, eq. (4.30), we need the first and second transport mean free paths for soft collisions, which are given by

$$\left(\lambda_{{\rm el},1}^{(s)}\right)^{-1} = \frac{2}{\lambda_{\rm el}} T_1(\mu_{\rm c}) \qquad \text{and} \qquad \left(\lambda_{{\rm el},2}^{(s)}\right)^{-1} = \frac{6}{\lambda_{\rm el}} \left[T_1(\mu_{\rm c}) - T_2(\mu_{\rm c})\right] \tag{4.38}$$

with

$$T_1(\mu_{\rm c}) = \int_0^{\mu_{\rm c}} \mu p_{\rm MW}(\mu)\, {\rm d}\mu \qquad \text{and} \qquad T_2(\mu_{\rm c}) = \int_0^{\mu_{\rm c}} \mu^2 p_{\rm MW}(\mu)\, {\rm d}\mu. \tag{4.39}$$

These latter quantities can be computed analytically as

$$T_1(\mu_c) = \begin{cases} (1-B)I_1(\mu_c) & \text{if } 0 \le \xi_c < \xi_0 \\ (1-B)I_1(\mu_0) + (\xi_c - \xi_0)\,\mu_0 & \text{if } \xi_0 \le \xi_c < \xi_0 + B \\ (1-B)I_1(\mu_c) + B\mu_0 & \text{if } \xi_0 + B \le \xi_c \le 1 \end{cases} \tag{4.40}$$

and

$$T_2(\mu_c) = \begin{cases} (1-B)I_2(\mu_c) & \text{if } 0 \le \xi_c < \xi_0 \\ (1-B)I_2(\mu_0) + (\xi_c - \xi_0)\,\mu_0^2 & \text{if } \xi_0 \le \xi_c < \xi_0 + B \\ (1-B)I_2(\mu_c) + B\mu_0^2 & \text{if } \xi_0 + B \le \xi_c \le 1 \end{cases} \tag{4.41}$$

with

$$I_1(\mu) \equiv A\left[(1+A)\ln\left(\frac{A+\mu}{A}\right) - \frac{(1+A)\mu}{A+\mu}\right] \tag{4.42}$$

and

$$I_2(\mu) \equiv A\left[\frac{(1+A)\mu^2}{A+\mu} - 2I_1(\mu)\right]. \tag{4.43}$$

The quantities ξ_0 and ξ_c are defined by eqs. (3.34) and (4.35), respectively.

4.2 Soft energy losses

The high-energy codes currently available implement different approximate methods to simulate inelastic collisions. Thus, ETRAN and ITS3 make use of the multiple scattering theories of Landau (1944) and Blunck and Leisegang (1950) to obtain the energy loss distribution due to inelastic collisions after a given path length; the production of secondary electrons is simulated by means of the Møller (1932) and Bhabha (1936) DCSs, which neglect binding effects. This approach accounts for the whole energy straggling, within the accuracy of the multiple scattering theory, but disregards the correlation between delta ray emission and energy loss in each track segment. As a consequence, energetic delta rays can be generated in a track segment where the energy lost by the primary particle is smaller than the energy of the emitted delta rays. EGS4 uses a mixed procedure to simulate collision energy losses: hard inelastic collisions are simulated from the Møller and Bhabha DCSs, thus neglecting binding effects, and soft inelastic collisions are described by means of the continuous slowing down approximation (CSDA), i.e. energy straggling due to soft inelastic collisions is ignored. As regards bremsstrahlung emission, EGS4 implements a mixed procedure in which hard radiative events are simulated in detail and use is made of the CSDA to simulate the effect of soft photon emission; ETRAN uses strictly detailed simulation.

To make the arguments more precise, we introduce the cutoff values W_{cc} and W_{cr}, and consider inelastic collisions with energy loss $W < W_{cc}$ and emission of bremsstrahlung photons with $W < W_{cr}$ as soft stopping interactions. The use of the CSDA to describe

soft interactions is well justified when the energy straggling due to these interactions is negligible, as happens when the cutoff energies W_{cc} and W_{cr} are both small, so that the fraction of the stopping power due to soft interactions is also small. To improve the description of energy straggling one should reduce the cutoff energies, but this enlarges the number of hard inelastic and radiative events to be simulated along each track and hence the simulation time. Our purpose is to go beyond the CSDA by introducing energy straggling in the description of soft stopping interactions. It is clear that, by proceeding in this way, we will be able to use larger values of the cutoff energies W_{cc} and W_{cr}, and hence speed up the simulation, without distorting the energy distributions.

In previous versions of PENELOPE, soft energy losses were simulated by using the mixed simulation algorithm described by Baró et al. (1995). The quantities that define the algorithm are the mean free paths $\lambda_{in}^{(h)}$ and $\lambda_{br}^{(h)}$ between hard collisions and hard radiative events, the stopping power S_s and the energy straggling parameter Ω_s^2 associated with soft interactions. These quantities are given by

$$\lambda_{in}^{(h)}(E) = \left(\mathcal{N} \int_{W_{cc}}^{E} \frac{d\sigma_{in}}{dW} \, dW \right)^{-1}, \tag{4.44}$$

$$\lambda_{br}^{(h)}(E) = \left(\mathcal{N} \int_{W_{cr}}^{E} \frac{d\sigma_{br}}{dW} \, dW \right)^{-1}, \tag{4.45}$$

$$S_s(E) = \mathcal{N} \int_0^{W_{cc}} W \frac{d\sigma_{in}}{dW} \, dW + \mathcal{N} \int_0^{W_{cr}} W \frac{d\sigma_{br}}{dW} \, dW \tag{4.46}$$

and

$$\Omega_s^2(E) = \mathcal{N} \int_0^{W_{cc}} W^2 \frac{d\sigma_{in}}{dW} \, dW + \mathcal{N} \int_0^{W_{cr}} W^2 \frac{d\sigma_{br}}{dW} \, dW. \tag{4.47}$$

To prevent $\lambda_{br}^{(h)}(E)$ from vanishing (infrared divergence), in PENELOPE the radiative cutoff energy W_{cr} is required to be larger than or equal to 10 eV.

Let us consider that a particle, electron or positron, travels a step of length s between two consecutive hard events of any kind (i.e. hard elastic or inelastic collisions, hard bremsstrahlung emissions, and annihilation in the case of positrons). Along this step, the particle is assumed to interact only through soft inelastic collisions and soft bremsstrahlung emission. We consider that the average energy loss in this path length, $S_s(E)s$, is much less than the initial energy E so that the DCSs can be assumed to stay essentially constant along the step. Let $G(s;\omega)$ denote the PDF of the energy loss ω along the path length s; this distribution satisfies the transport equation (Landau, 1944)

$$\frac{\partial G(s;\omega)}{\partial s} = \mathcal{N} \int_0^{\infty} [G(s;\omega - W) - G(s;\omega)] \, \sigma_s(E;W) \, dW \tag{4.48}$$

with the initial value $G(0;\omega) = \delta(\omega)$. Here, $\sigma_s(E;W)$ stands for the DCS for soft stopping interactions, i.e.

$$\sigma_s(E;W) \equiv \frac{d\sigma_s}{dW} = \frac{d\sigma_{in}}{dW} \Theta(W_{cc} - W) + \frac{d\sigma_{br}}{dW} \Theta(W_{cr} - W), \tag{4.49}$$

where $\Theta(x)$ is the step function. A closed formal solution of the integral equation (4.48) may be obtained by considering its Fourier, or Laplace, transform with respect to ω (see e.g. Landau, 1944, Blunck and Leisegang, 1950). For our purposes it is only necessary to know the first moments of the energy loss distribution after the path length s,

$$\langle\omega^n\rangle \equiv \int_0^\infty \omega^n G(s;\omega)\,\mathrm{d}\omega. \tag{4.50}$$

From eq. (4.48) it follows that

$$\begin{aligned}\frac{\mathrm{d}}{\mathrm{d}s}\langle\omega^n\rangle &= \mathcal{N}\int_0^\infty \mathrm{d}\omega \int_0^\infty \mathrm{d}W\,\omega^n\left[G(s;\omega-W)-G(s;\omega)\right]\sigma_\mathrm{s}(E;W)\\ &= \mathcal{N}\left(\int_0^\infty \mathrm{d}\omega' \int_0^\infty \mathrm{d}W\,(\omega'+W)^n G(s;\omega')\sigma_\mathrm{s}(E;W) - \langle\omega^n\rangle\int_0^\infty \sigma_\mathrm{s}(E;W)\,\mathrm{d}W\right)\\ &= \sum_{k=1}^{n}\frac{n!}{k!(n-k)!}\langle\omega^{n-k}\rangle\mathcal{N}\int_0^\infty W^k\sigma_\mathrm{s}(E;W)\,\mathrm{d}W,\end{aligned} \tag{4.51}$$

where use has been made of the fact that $\sigma_\mathrm{s}(E;W)$ vanishes when $W<0$. In particular, we have

$$\frac{\mathrm{d}}{\mathrm{d}s}\langle\omega\rangle = \mathcal{N}\int_0^\infty W\sigma_\mathrm{s}(E;W)\,\mathrm{d}W = S_\mathrm{s}, \tag{4.52}$$

$$\begin{aligned}\frac{\mathrm{d}}{\mathrm{d}s}\langle\omega^2\rangle &= 2\langle\omega\rangle\mathcal{N}\int_0^\infty W\sigma_\mathrm{s}(E;W)\,\mathrm{d}W + \mathcal{N}\int_0^\infty W^2\sigma_\mathrm{s}(E;W)\,\mathrm{d}W\\ &= 2\langle\omega\rangle S_\mathrm{s} + \Omega_\mathrm{s}^2\end{aligned} \tag{4.53}$$

and, hence,

$$\langle\omega\rangle = S_\mathrm{s}s, \tag{4.54}$$

$$\langle\omega^2\rangle = (S_\mathrm{s}s)^2 + \Omega_\mathrm{s}^2 s. \tag{4.55}$$

The variance of the energy loss distribution is

$$\mathrm{var}(\omega) = \langle\omega^2\rangle - \langle\omega\rangle^2 = \Omega_\mathrm{s}^2 s, \tag{4.56}$$

i.e. the energy straggling parameter Ω_s^2 equals the variance increase per unit path length.

The key point in our argument is that soft interactions involve only comparatively small energy losses. If the number of soft interactions along the path length s is statistically sufficient, it follows from the central limit theorem that the energy loss distribution is Gaussian with mean $S_\mathrm{s}s$ and variance $\Omega_\mathrm{s}^2 s$, i.e.

$$G(s;\omega) \simeq \frac{1}{(2\pi\Omega_\mathrm{s}^2(E)s)^{1/2}}\exp\left[-\frac{(\omega-S_\mathrm{s}(E)s)^2}{2\Omega_\mathrm{s}^2(E)s}\right]. \tag{4.57}$$

This result is accurate only if 1) the average energy loss $S_\mathrm{s}(E)s$ is much smaller than E (so that the DCS $\mathrm{d}\sigma_\mathrm{s}/\mathrm{d}W$ is nearly constant along the step) and 2) its standard

deviation $[\Omega_s^2(E)s]^{1/2}$ is much smaller than its mean $S_s(E)s$ (otherwise there would be a finite probability of negative energy losses), i.e.

$$\left[\Omega_s^2(E)s\right]^{1/2} \ll S_s(E)s \ll E. \tag{4.58}$$

Requirement 1) implies that the cutoff energies W_{cc} and W_{cr} for delta ray production and photon emission have to be relatively small. The second requirement holds for path lengths larger than $s_{crit} = \Omega_s^2/S_s^2$.

Now, we address ourselves to the problem of simulating the energy losses due to soft stopping interactions between two consecutive hard events. The distribution (4.57) gives the desired result when conditions (4.58) are satisfied. In fact, the use of a Gaussian distribution to simulate the effect of soft stopping interactions was previously proposed by Andreo and Brahme (1984). Unfortunately, the step lengths found in our simulations are frequently too short for conditions (4.58) to hold (i.e. s is usually less than s_{crit}). To get over this problem, we replace the actual energy loss distribution $G(s;\omega)$ by a simpler "equivalent" distribution $G_a(s;\omega)$ with the same mean and variance, given by eqs. (4.54) and (4.56). Other details of the adopted distribution have no effect on the simulation results, provided that the number of steps along each track is statistically sufficient (say, larger than ~ 20). PENELOPE generates ω from the following distributions

• Case I. If $\langle\omega\rangle^2 > 9\,\mathrm{var}(\omega)$, we use a truncated Gaussian distribution,

$$G_{a,I}(s;\omega) = \begin{cases} \exp\left[-\dfrac{(\omega-\langle\omega\rangle)^2}{2(1.015387\sigma)^2}\right] & \text{if } |\omega-\langle\omega\rangle| < 3\,\sigma. \\ 0 & \text{otherwise.} \end{cases} \tag{4.59}$$

where $\sigma = [\mathrm{var}(\omega)]^{1/2}$ is the standard deviation and the numerical factor 1.015387 corrects for the effect of the truncation. Notice that the shape of this distribution is very similar to that of the "true" energy-loss distribution, eq. (4.57). Random sampling from (4.59) is performed by means of the Box-Müller method, eq. (1.54), rejecting the generated ω's that are outside the interval $\langle\omega\rangle \pm 3\sigma$.

• Case II. When $3\,\mathrm{var}(\omega) < \langle\omega\rangle^2 < 9\,\mathrm{var}(\omega)$, the energy loss is sampled from the uniform distribution

$$G_{a,II}(s;\omega) = U_{\omega_1,\omega_2}(\omega) \tag{4.60}$$

with

$$\omega_1 = \langle\omega\rangle - \sqrt{3}\,\sigma, \qquad \omega_2 = \langle\omega\rangle + \sqrt{3}\,\sigma. \tag{4.61}$$

• Case III. Finally, when $\langle\omega\rangle^2 < 3\,\mathrm{var}(\omega)$, the adopted distribution is an admixture of a delta and a uniform distribution,

$$G_{a,III}(s;\omega) = a\delta(\omega) + (1-a)U_{0,\omega_0}(\omega) \tag{4.62}$$

with

$$a = \frac{3\mathrm{var}(\omega) - \langle\omega\rangle^2}{3\mathrm{var}(\omega) + 3\langle\omega\rangle^2} \quad \text{and} \quad \omega_0 = \frac{3\mathrm{var}(\omega) + 3\langle\omega\rangle^2}{2\langle\omega\rangle}. \tag{4.63}$$

It can be easily verified that these distributions have the required mean and variance. It is also worth noticing that they yield ω values that are less than

$$\omega_{\max} = \begin{cases} \langle\omega\rangle + 3\sigma & \text{in case I,} \\ \omega_2 & \text{in case II,} \\ \omega_0 & \text{in case III.} \end{cases} \tag{4.64}$$

$\omega_{\max}$ is normally much less than the kinetic energy E of the transported particle. Energy losses larger than E might be generated only when the step length s has a value of the order of the Bethe range, but this never happens in practical simulation (see below). It is worth noticing that, after a moderately large number of steps, this simple simulation scheme effectively yields an energy loss distribution that has the correct first and second moments and is similar in shape to the "true" distribution. Further improvements of the distribution of soft energy losses would mean considering higher order moments of the single scattering inelastic DCS given by eq. (4.49).

In spatial-dose calculations, the energy loss ω due to soft stopping interactions can be considered to be locally deposited at a random position uniformly distributed along the step. This procedure yields dose distributions identical to those obtained by assuming that the energy loss is deposited at a constant rate along the step, but is computationally simpler. According to this, PENELOPE simulates the combined effect of all soft elastic collisions and soft stopping interactions that occur between a pair of successive hard events, separated a distance s, as a single event (a hinge) in which the particle changes its direction of movement according to the distribution $F_{\rm a}(s;\mu)$, eqs. (4.30)-(4.32), and loses energy ω that is generated from the distribution $G_{\rm a}(s;\omega)$, eqs. (4.59)-(4.63). The position of the hinge is sampled uniformly along the step, as in the case of purely elastic scattering (section 4.1.2). When the step crosses an interface (see fig. 4.2), the artificial event is simulated only when its position lies in the initial material; otherwise the track is stopped at the interface and restarted in the new material. It can be easily verified that the particle reaches the interface not only with the correct average direction of movement, but also with the correct average energy, $E - S_{\rm s}t$.

4.2.1 Energy dependence of the soft DCS

The simulation model for soft energy losses described above is based on the assumption that the associated energy-loss DCS does not vary with the energy of the transported particle. To account for the energy dependence of the DCS in a rigorous way, we have to start from the transport equation [cf. eq. (4.48)]

$$\begin{aligned} \frac{\partial G(s;\omega)}{\partial s} &= \mathcal{N}\int_0^\infty G(s;\omega-W)\,\sigma_{\rm s}(E_0-\omega+W;W)\,{\rm d}W \\ &\quad - \mathcal{N}\int_0^\infty G(s;\omega)\,\sigma_{\rm s}(E_0-\omega;W)\,{\rm d}W, \end{aligned} \tag{4.65}$$

where E_0 denotes the kinetic energy of the particle at the beginning of the step. We desire to obtain expressions for the first and second moments, $\langle\omega\rangle$ and $\langle\omega^2\rangle$, of the multiple

scattering energy-loss distribution, which define the artificial distribution $G_a(s;\omega)$ as described above. Unfortunately, for a realistic DCS, these moments can only be obtained after arduous numerical calculations and we have to rely on simple approximations that can be easily implemented in the simulation code.

Let us consider that, at least for relatively small fractional energy losses, the DCS varies linearly with the kinetic energy of the particle,

$$\sigma_s(E_0-\omega;W) \simeq \sigma_s(E_0;W) - \left[\frac{\partial\sigma_s(E;W)}{\partial E}\right]_{E=E_0}\omega. \tag{4.66}$$

We recall that we are considering only soft energy-loss interactions (inelastic collisions and bremsstrahlung emission) for which the cutoff energies, W_{cc} and W_{cr}, do not vary with E. Therefore, the upper limit of the integrals in the right hand side of eq. (4.65) is finite and independent of the energy of the particle. The stopping power $S_s(E_0-\omega)$ can then be approximated as

$$S_s(E_0-\omega) \equiv \mathcal{N}\int W\sigma_s(E_0-\omega;W)\,\mathrm{d}W \simeq S_s(E_0) - S_s'(E_0)\,\omega, \tag{4.67}$$

where the prime denotes the derivative with respect to E. Similarly, for the straggling parameter $\Omega_s^2(E)$ we have

$$\Omega_s^2(E_0-\omega) \equiv \mathcal{N}\int W^2\sigma_s(E_0-\omega;W)\,\mathrm{d}W \simeq \Omega_s^2(E_0) - \Omega_s^{2\prime}(E_0)\,\omega. \tag{4.68}$$

From eq. (4.65) it follows that the moments of the multiple scattering distribution,

$$\langle\omega^n\rangle = \int \omega^n G(s;\omega)\,\mathrm{d}\omega,$$

satisfy the equations

$$\begin{aligned}\frac{\mathrm{d}}{\mathrm{d}s}\langle\omega^n\rangle &= \mathcal{N}\int \mathrm{d}w \int \mathrm{d}W\ [(w+W)^n G(s;w)\sigma_s(E_0-w;W)] \\ &\quad - \mathcal{N}\int \mathrm{d}\omega \int \mathrm{d}W\ \omega^n G(s;\omega)\sigma_s(E_0-\omega;W) \\ &= \mathcal{N}\sum_{k=1}^{n}\frac{n!}{k!(n-k)!}\int \mathrm{d}\omega \int \mathrm{d}W\,\omega^{n-k}W^k G(s;\omega)\sigma_s(E_0-\omega;W).\end{aligned} \tag{4.69}$$

By inserting the approximation (4.66), we obtain

$$\frac{\mathrm{d}}{\mathrm{d}s}\langle\omega^n\rangle = \sum_{k=1}^{n}\frac{n!}{k!(n-k)!}\left(\left\langle\omega^{n-k}\right\rangle M_k - \left\langle\omega^{n-k+1}\right\rangle M_k'\right), \tag{4.70}$$

where

$$M_k \equiv \mathcal{N}\int W^k\sigma_s(E_0;W)\,\mathrm{d}W \tag{4.71}$$

and

$$M_k' \equiv \mathcal{N} \int W^k \left[\frac{\partial \sigma_s(E;W)}{\partial E}\right]_{E=E_0} \mathrm{d}W = \left[\frac{\mathrm{d}M_k}{\mathrm{d}E}\right]_{E=E_0}. \tag{4.72}$$

The equations (4.70) with the boundary conditions $\langle\omega^n\rangle_{s=0} = 0$ can now be solved sequentially to any order. For $n = 1$ we have

$$\frac{\mathrm{d}}{\mathrm{d}s}\langle\omega\rangle = S_s(E_0) - S_s'(E_0)\langle\omega\rangle, \tag{4.73}$$

which yields

$$\langle\omega\rangle = \frac{S_s(E_0)}{S_s'(E_0)}\left\{1 - \exp\left[-S_s'(E_0)s\right]\right\}. \tag{4.74}$$

The equation for $n = 2$ reads,

$$\frac{\mathrm{d}}{\mathrm{d}s}\langle\omega^2\rangle = \Omega_s^2(E_0) + \left[2S_s(E_0) - \Omega_s^{2\prime}(E_0)\right]\langle\omega\rangle - 2S_s'(E_0)\langle\omega^2\rangle, \tag{4.75}$$

and its solution is

$$\begin{aligned}\langle\omega^2\rangle &= \Omega_s^2(E_0)\frac{1-\exp[-2S_s'(E_0)s]}{2S_s'(E_0)} \\ &\quad + s\left[2S_s(E_0) - \Omega_s^{2\prime}(E_0)\right]S_s(E_0)\left[\frac{1-\exp[-S_s'(E_0)s]}{2S_s'(E_0)}\right]^2.\end{aligned} \tag{4.76}$$

Hence,

$$\begin{aligned}\mathrm{var}(\omega) &= \langle\omega^2\rangle - \langle\omega\rangle^2 \\ &= \Omega_s^2(E_0)\frac{1-\exp[-2S_s'(E_0)s]}{2S_s'(E_0)} - 2\Omega_s^{2\prime}(E_0)S_s(E_0)\left[\frac{1-\exp[-S_s'(E_0)s]}{2S_s'(E_0)}\right]^2.\end{aligned} \tag{4.77}$$

Since these expressions are derived from the linear approximation, eq. (4.66), it is consistent to evaluate $\langle\omega\rangle$ and $\mathrm{var}(\omega)$ from their Taylor expansions to second order,

$$\begin{aligned}\langle\omega\rangle &= S_s(E_0)\,s\left[1 - \frac{1}{2}S_s'(E_0)\,s + \mathcal{O}(s^2)\right] \\ &\simeq S_s(E_0)\,s\left\{1 - \frac{1}{2}\left[\frac{\mathrm{d}\ln S_s(E)}{\mathrm{d}E}\right]_{E=E_0} S_s(E_0)\,s\right\}\end{aligned} \tag{4.78}$$

and

$$\begin{aligned}\mathrm{var}(\omega) &= \Omega_s^2(E_0)\,s - \left[\frac{1}{2}\Omega_s^{2\prime}(E_0)\,S_s(E_0) + \Omega_s^2(E_0)\,S_s'(E_0)\right]s^2 + \mathcal{O}(s^3) \\ &\simeq \Omega_s^2(0)\,s\left\{1 - \left[\frac{1}{2}\frac{\mathrm{d}\ln\Omega_s^2(E)}{\mathrm{d}E} + \frac{\mathrm{d}\ln S_s(E)}{\mathrm{d}E}\right]_{E=E_0} S_s(E_0)\,s\right\},\end{aligned} \tag{4.79}$$

 3569

where the logarithmic derivatives have been introduced for numerical convenience. The factors in curly brackets account for the global effect of the energy dependence of the soft energy-loss DCS (within the linear approximation). To simulate soft energy losses, we sample ω from the artificial distribution $G_a(\omega; s)$, eqs. (4.59) to (4.63), with the "correct" first moment and variance, given by expressions (4.78) and (4.79). In PENELOPE, we use step lengths s such that the fractional energy loss along each step is relatively small (see below) and, consequently, the energy-dependence correction is also small (i.e. the correcting factors are close to unity).

4.3 Combined scattering and energy loss

Up to this point, soft scattering and energy loss have been regarded as essentially independent processes, while in reality they coexist. In this section, we consider their interplay and set the basis of an algorithm that simulates their combined effect.

Ours is a mixed algorithm, where hard simulations are described individually from the associated DCSs (see chapter 3). These interactions are 1) hard elastic collisions, "el", 2) hard inelastic collisions, "in", 3) hard bremsstrahlung photon emission "br", and, in the case of positrons, 4) positron annihilation, "an". The mean free path between consecutive hard events, $\lambda_T^{(h)}$, is given by

$$\left[\lambda_T^{(h)}\right]^{-1} = \mathcal{N}\sigma_T^{(h)} = \mathcal{N}\left[\sigma_{el}^{(h)} + \sigma_{in}^{(h)} + \sigma_{br}^{(h)}\left(+\sigma_{an}^{(h)}\right)\right] \equiv \Sigma_h, \tag{4.80}$$

where $\sigma_T^{(h)}$ is the total atomic cross section for hard interactions. We recall that the inverse mean free path, Σ_h, gives the interaction probability per unit path length. In the absence of soft energy-loss events, the PDF of the step length s between two successive hard events (or from a given point in the track to the next hard event) is

$$p(s) = \Sigma_h \exp\left(-\Sigma_h s\right). \tag{4.81}$$

In each hard event, one and only one interaction (i="el", "in", "br" or "an") occurs with probability

$$p_i = \sigma_i^{(h)}/\sigma_T^{(h)}. \tag{4.82}$$

When soft energy-losses are considered, the PDF of the distance s travelled by the particle to the following hard interaction is not given by eq. (4.81), because the mean free path $\lambda_T^{(h)}$ varies with energy and may change appreciably along a single step. The simplest way to cope with this problem is to limit the length of the step to make sure that the *average* energy loss is much smaller than the kinetic energy E at the beginning of the step, and consider that $\lambda_T^{(h)}(E)$ remains essentially constant along the step. Then, the mean energy loss in a step is given by

$$\langle \Delta E \rangle = \lambda_T^{(h)} S(E), \tag{4.83}$$

where

$$S(E) = S_{\text{in}}(E) + S_{\text{br}}(E) \tag{4.84}$$

is the total stopping power. Since the mean free path between consecutive hard events of any kind is shorter than the mean free path between hard elastic events, the energy loss per step can be limited by re-defining the hard mean free path. If we wish to tolerate average fractional energy losses $\Delta E/E$ along a step of the order of C_2 (a small value, say, 0.05), we simply take

$$\lambda_{\text{el}}^{(\text{h})}(E) = \max\left\{\lambda_{\text{el}}(E), \min\left[C_1\lambda_{\text{el},1}(E), C_2\frac{E}{S(E)}\right]\right\}. \tag{4.85}$$

This effectively limits the average energy loss per step at the expense of increasing the frequency of hard elastic events. The parameters C_1 and C_2 in eq. (4.85), to be selected by the user, determine the computer time needed to simulate each track. Ideally, they should not have any influence on the *accuracy* of the simulation results. This happens only when their values are sufficiently small (see below).

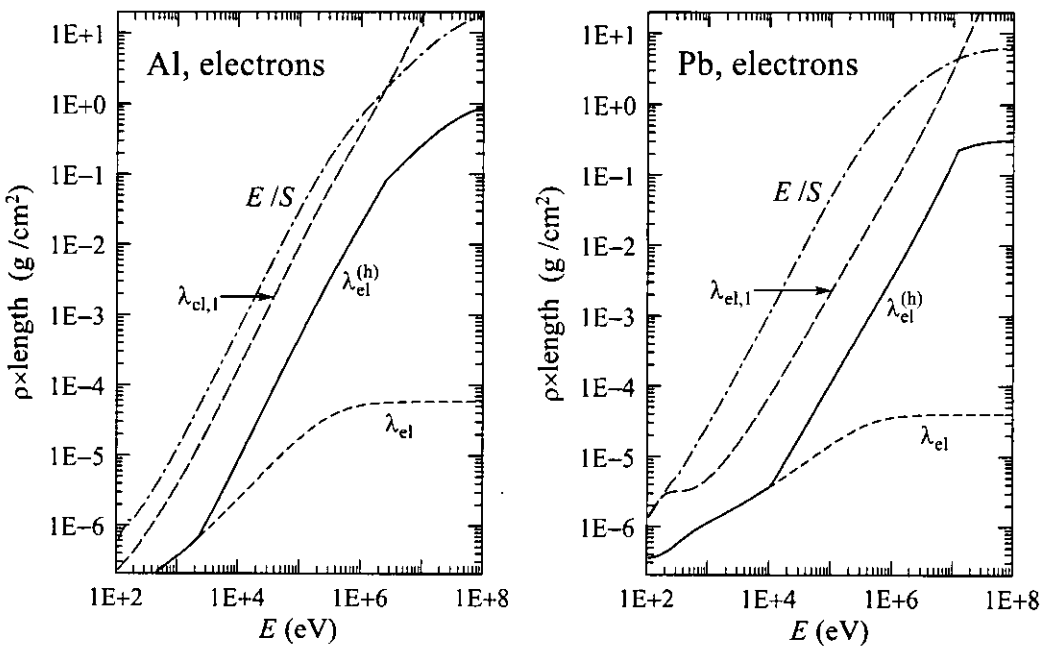


Figure 4.3: Elastic mean free path λ_{el}, first transport mean free path $\lambda_{\text{el},1}$ and $E/S(E)$ for electrons in aluminium and lead. The solid line represents the mean free path between hard elastic events $\lambda_{\text{el}}^{(\text{h})}$ obtained from eq. (4.85) with $C_1 = C_2 = 0.05$.

It should be noted that C_1 and C_2 act on different energy domains. This is illustrated in fig. 4.3, where the lengths λ_{el}, $\lambda_{\text{el},1}$ and E/S for electrons in aluminium and lead are represented as functions of the kinetic energy. The mean free path $\lambda_{\text{el}}^{(\text{h})}$ for hard elastic

events, determined from the prescription (4.85) with $C_1 = C_2 = 0.05$ is also plotted. For low energies, $\lambda_{\rm el}^{(\rm h)} = \lambda_{\rm el}$ and the simulation is purely detailed ($\mu_{\rm c} = 0$). For intermediate energies, $\lambda_{\rm el}^{(\rm h)} = C_1 \lambda_{\rm el,1}$, whereas $\lambda_{\rm el}^{(\rm h)} = C_2 E/S(E)$ in the high-energy domain. From fig. 4.3 it is clear that increasing the value of C_2 does not have any effect on the simulation of electron tracks with initial energies that are less than $\sim$ 10 MeV.

4.3.1 Variation of $\lambda_{\rm T}^{(\rm h)}$ with energy

With the definition (4.85) of the hard elastic mean free path, we only set a limit on the *average* step length. However, since s is sampled from the exponential distribution, its realizations fluctuate amply about the average value. On the other hand, the soft energy loss ω along a step of given length s also fluctuates about the mean value $\langle\omega\rangle$ given by eq. (4.78). This means that the inverse mean free path $\Sigma_{\rm h}(E)$ varies along the step in an essentially unpredictable way.

Let us consider for a moment that the CSDA is applicable (i.e. that the effect of soft energy straggling is negligible). In this case, there is a one-by-one correspondence between the kinetic energy E of the electron and the travelled path length s,

$$s = \int_E^{E_0} \frac{{\rm d}E'}{S_{\rm s}(E')}, \tag{4.86}$$

where $S_{\rm s}(E)$ is the soft stopping power, eq. (4.46) [we consider that no hard interactions occur along the step]. Equivalently,

$$\frac{{\rm d}s}{{\rm d}E} = -\frac{1}{S_{\rm s}(E)}. \tag{4.87}$$

Thus, the inverse mean free path $\Sigma_{\rm h}$ can be formally considered as a function of the path length s. The probability $p(s)\,{\rm d}s$ of having the first hard interaction when the particle has travelled a length in the interval $(s, s + {\rm d}s)$ is determined by the equation [cf. eq. (1.72)]

$$p(s) = \Sigma_{\rm h}(s) \int_s^\infty p(s')\,{\rm d}s', \tag{4.88}$$

with the normalization condition,

$$\int_0^\infty p(s)\,{\rm d}s = 1. \tag{4.89}$$

Instead of the path length s, it is convenient to consider the dimensionless variable

$$q \equiv \int_E^{E_0} \frac{\Sigma_{\rm h}(E')}{S_{\rm s}(E')} {\rm d}E' = \int_0^s \Sigma_{\rm h}(s')\,{\rm d}s', \tag{4.90}$$

which varies with energy and

$$\frac{{\rm d}q}{{\rm d}E} = -\frac{\Sigma_{\rm h}(E)}{S_{\rm s}(E)}. \tag{4.91}$$

The PDF of q is

$$\pi(q) = p(s)\frac{\mathrm{d}s}{\mathrm{d}q} = p(s)\,\frac{\mathrm{d}s}{\mathrm{d}E}\,\frac{\mathrm{d}E}{\mathrm{d}q} = p(s)\frac{1}{\Sigma_{\mathrm{h}}(s)}. \tag{4.92}$$

From eq. (4.88) it follows that $\pi(q)$ satisfies the equation

$$\pi(q) = \int_q^\infty \pi(q')\,\mathrm{d}q'. \tag{4.93}$$

Therefore, q is distributed exponentially,

$$\pi(q) = \exp(-q). \tag{4.94}$$

The PDF of the step length s is obtained by inverting the transformation (4.90),

$$p(s) = \Sigma_{\mathrm{h}}(s)\exp\left(-\int_0^s \Sigma_{\mathrm{h}}(s')\,\mathrm{d}s'\right). \tag{4.95}$$

It is not practical to sample s from this complicated PDF. It is much more convenient to sample q [as $-\ln\xi$, cf. eq. (1.36)] and then determine s from (4.90), which can be inverted numerically (for practical details, see Berger, 1998). Although this sampling method effectively accounts for the energy dependence of $\Sigma_{\mathrm{s}}(E)$, it is applicable only to simulations in the CSDA.

A more versatile algorithm for sampling the position of hard events, still within the CSDA, is the following. We let the electron move in steps of maximum length s_{max}, a value specified by the user. This determines the maximum energy loss along the step,

$$\omega_{\mathrm{max}} = \int_0^{s_{\mathrm{max}}} S_{\mathrm{s}}(s)\,\mathrm{d}s. \tag{4.96}$$

Let $\Sigma_{\mathrm{h,max}}$ denote an upper bound for the inverse mean free path of hard events in the swept energy interval, i.e.

$$\Sigma_{\mathrm{h,max}} > \max\{\Sigma_{\mathrm{h}}(E), E \in (E_0 - \omega_{\mathrm{max}}, E_0)\} \tag{4.97}$$

We now assume that the electron may undergo fictitious events in which the energy and direction remain unaltered (delta interactions). The inverse mean free path of these interactions is defined as

$$\Sigma_\delta(E) = \Sigma_{\mathrm{h,max}} - \Sigma_{\mathrm{h}}(E), \tag{4.98}$$

so that the inverse mean free path of the combined process (delta interactions + hard events) equals $\Sigma_{\mathrm{h,max}}$, a constant. Owing to the Markovian character of the processes, the introduction of delta interactions does not influence the path-length distribution between hard events. Therefore, the occurrence of hard events can be sampled by means of the following simple algorithm,

(i) Sample a distance s from the exponential distribution with inverse mean free path $\Sigma_{\mathrm{h,max}}$, i.e. $s = (-\ln\xi)/\Sigma_{\mathrm{h,max}}$.

(ii) If $s > s_{max}$, move the electron a path length s_{max} and determine the soft energy loss ω along this path length. Modify the electron energy[2], $E \leftarrow E - \omega$, and assume that a delta interaction occurs at the end of the step.

(iii) If $s < s_{max}$, move the electron a step of length s. Determine the energy loss ω and update the energy, $E \leftarrow E - \omega$. Sample a random number ξ.

(1) If $\xi\Sigma_{h,max} < \Sigma_h(E)$, simulate a hard interaction

(2) Otherwise, assume that the particle undergoes a delta interaction.

(iv) Return to (i).

It is clear that the path-length s to the first hard interaction generated with this algorithm follows the PDF (4.95). The interesting peculiarity of this algorithm is that it makes no explicit reference to the CSDA. Therefore, it can be adopted in mixed simulations with soft-energy-loss straggling, provided only that an upper bound exists for the energy ω lost along the path length s_{max}.

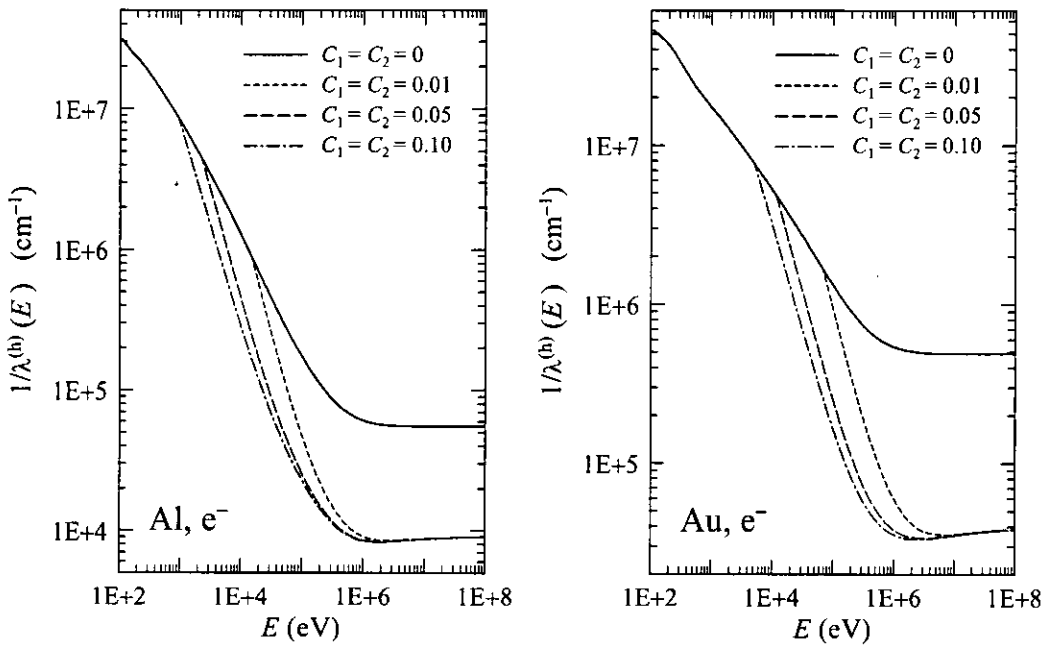


Figure 4.4: Inverse mean free path (interaction probability per unit path length) for hard interactions of electrons in Al and Au for the indicated values of the simulation parameters. The plotted curves were calculated with $W_{cc} = W_{cr} = 100$ eV.

Fortunately, the energy loss generated from the artificial distribution $G_a(\omega; s)$, eqs. (4.59)-(4.63), is always less than ω_{max}, eq. (4.64). Indeed, in case I we use the truncated

[2]In the description of the algorithms we use the symbol $\leftarrow$ in expressions such as "$a \leftarrow b$" to indicate that the value b replaces the value of a.

Gaussian distribution (4.59) just to enforce this property. In our mixed simulation we shall select a maximum step length $s_{\max}$, which serves to set an upper bound for the energy that the transported electrons may lose along each step. Since the hard inverse mean free path $\Sigma_h(E)$ has a broad minimum (and no local maxima) in the whole energy interval of interest (see fig. 4.4), the maximum value of Σ_h within a certain energy interval (E_1, E_2) occurs at one of the end points. This makes the practical implementation of the above algorithm very easy.

4.3.2 Scattering by atomic electrons

Most of the existing high-energy simulation codes have difficulties in accounting for the angular deflections of the projectile due to inelastic collisions (see e.g. Jenkins et al., 1988). The inelastic cross section differential in the scattering angle can be calculated approximately in terms of the incoherent scattering function (see e.g. Mott and Massey, 1965). This was the approach followed by Fano (1954) in order to introduce electron scattering effects in the Molière (1948) multiple scattering theory. However, the DCS calculated in this way accounts for all excitations and, hence, it is not adequate for mixed simulations, where the part of electron scattering due to hard collisions is explicitly simulated. Moreover, the calculation of the DCS from the incoherent scattering function involves an average over excitation energies that cannot be performed exactly; instead an effective "minimum momentum transfer" is introduced, which must be estimated empirically. This may cause inconsistencies for low-energy projectiles. A more consistent approach (Baró et al., 1995) is obtained by simply computing the restricted angular DCS, for soft collisions with $W < W_{cc}$, from our inelastic scattering model (see section 3.2), as follows.

We recall that the recoil energy Q is given by (see appendix B)

$$Q(Q + 2m_e c^2) = c^2(p^2 + p'^2 - 2pp' \cos\theta), \tag{4.99}$$

where p and p' are the magnitudes of the momentum of the projectile before and after the collision,

$$(cp)^2 = E(E + 2m_e c^2) \qquad \text{and} \qquad (cp')^2 = (E - W)(E - W + 2m_e c^2). \tag{4.100}$$

In soft distant interactions, the angular deflection $\mu = (1 - \cos\theta)/2$ and the recoil energy Q are related through

$$Q(Q + 2m_e c^2) = 4cp\, cp_k \mu + (cp - cp_k)^2, \tag{4.101}$$

where p_k is the momentum of the projectile after the collision,

$$(cp_k)^2 = (E - W_k)(E - W_k + 2m_e c^2). \tag{4.102}$$

The cross section for soft distant interactions[3], eq. (3.60), can then be expressed in terms

[3]Distant transverse interactions do not cause scattering.

of the variable μ as

$$\frac{\mathrm{d}\sigma_{\mathrm{dis,l}}}{\mathrm{d}\mu} = \frac{4\pi e^4}{\mathrm{m_e}v^2}\sum_k f_k \frac{1}{W_k}\,\frac{\mathrm{m_e}c^2}{4\,cp\,cp_k\mu + (cp - cp_k)^2}\,\frac{4\,cp\,cp_k}{2(Q + \mathrm{m_e}c^2)}. \tag{4.103}$$

Considering that $Q \ll \mathrm{m_e}c^2$ for the majority of soft distant collisions, we have

$$\frac{\mathrm{d}\sigma_{\mathrm{dis,l}}}{\mathrm{d}\mu} = \frac{2\pi e^4}{\mathrm{m_e}v^2}\sum_k f_k \frac{1}{W_k}\,\frac{1}{R_k + \mu}, \qquad 0 < \mu < \mu_k, \tag{4.104}$$

where

$$R_k = \frac{(cp - cp_k)^2}{4\,cp\,cp_k} \tag{4.105}$$

and

$$\mu_k = \mu(Q = W_k) = \frac{W_k(W_k + 2\mathrm{m_e}c^2) - (cp - cp_k)^2}{4cp\,cp_k}. \tag{4.106}$$

On the other hand, the DCS per unit oscillator strength for soft ($W < W_{\mathrm{cc}}$) close collisions with the i-th oscillator is given by [see eqs. (3.71) and (3.77)]

$$\frac{\mathrm{d}\sigma_{\mathrm{clo}}^{(\pm)}}{\mathrm{d}W} = \frac{2\pi e^4}{\mathrm{m_e}v^2}\sum_k f_k \frac{1}{W^2}F^{(\pm)}(E, W). \tag{4.107}$$

The angular deflection and the energy loss are related by (3.118), which implies that

$$W = \frac{E(E + 2\mathrm{m_e}c^2)2(\mu - \mu^2)}{2E(\mu - \mu^2) + \mathrm{m_e}c^2} \tag{4.108}$$

and

$$\frac{\mathrm{d}W}{\mathrm{d}\mu} = \frac{E(E + 2\mathrm{m_e}c^2)\mathrm{m_e}c^2 2(1 - 2\mu)}{[2E(\mu - \mu^2) + \mathrm{m_e}c^2]^2}. \tag{4.109}$$

Therefore,

$$\frac{\mathrm{d}\sigma_{\mathrm{clo}}^{(\pm)}}{\mathrm{d}\mu} = \frac{2\pi e^4}{\mathrm{m_e}v^2}\sum_k f_k \frac{1}{W^2}F^{(\pm)}(E, W)\frac{\mathrm{d}W}{\mathrm{d}\mu}, \qquad \mu_k < \mu < \mu_{\mathrm{cc}}, \tag{4.110}$$

where

$$\mu_{\mathrm{cc}} = \mu(Q = W_{\mathrm{cc}}) = \frac{W_{\mathrm{cc}}(W_{\mathrm{cc}} + 2\mathrm{m_e}c^2) - (cp - cp_{\mathrm{cc}})^2}{4cp\,cp_{\mathrm{cc}}} \tag{4.111}$$

with

$$(cp_{\mathrm{cc}})^2 = (E - W_{\mathrm{cc}})(E - W_{\mathrm{cc}} + 2\mathrm{m_e}c^2). \tag{4.112}$$

The angular DCS for soft inelastic interactions is then given by

$$\begin{aligned}\frac{\mathrm{d}\sigma_{\mathrm{s}}}{\mathrm{d}\mu} &= \frac{\mathrm{d}\sigma_{\mathrm{dis,l}}}{\mathrm{d}\mu} + \frac{\mathrm{d}\sigma_{\mathrm{clo}}^{(\pm)}}{\mathrm{d}\mu} \\ &= \frac{2\pi e^4}{\mathrm{m_e}v^2}\sum_k f_k \left\{\frac{1}{W_k}\,\frac{1}{R_k + \mu} + \frac{1}{W^2}F^{(\pm)}(E, W)\frac{\mathrm{d}W}{\mathrm{d}\mu}\right\},\end{aligned} \tag{4.113}$$

where the summations extend over the oscillators with resonance energy less than W_{cc} and greater than W_{max}, and each term contributes only for the μ-intervals indicated above. The mean free path and the first and second transport mean free paths for soft inelastic scattering are

$$\left[\lambda_{in}^{(s)}\right]^{-1} = \mathcal{N}\int_0^{\mu_2} \frac{d\sigma_{in}^{(s)}}{d\mu}\, d\mu, \tag{4.114}$$

$$\left[\lambda_{in,1}^{(s)}\right]^{-1} = \mathcal{N}\int_0^{\mu_2} 2\mu\frac{d\sigma_{in}^{(s)}}{d\mu}\, d\mu \tag{4.115}$$

and

$$\left[\lambda_{in,2}^{(s)}\right]^{-1} = \mathcal{N}\int_0^{\mu_2} 6(\mu-\mu^2)\frac{d\sigma_{in}^{(s)}}{d\mu}\, d\mu. \tag{4.116}$$

In PENELOPE, soft electronic scattering is simulated together with soft elastic scattering, by means of the artificial distribution (4.30). The combined process is described by the transport mean free paths

$$\left[\lambda_{comb,1}^{(s)}\right]^{-1} = \left[\lambda_{el,1}^{(s)}\right]^{-1} + \left[\lambda_{in,1}^{(s)}\right]^{-1} \tag{4.117}$$

and

$$\left[\lambda_{comb,2}^{(s)}\right]^{-1} = \left[\lambda_{el,2}^{(s)}\right]^{-1} + \left[\lambda_{in,2}^{(s)}\right]^{-1}. \tag{4.118}$$

Thus, to account for soft electronic scattering we only have to replace the soft elastic transport mean free paths by those of the combined process.

4.3.3 Bielajew's alternate random hinge

Angular deflections due to soft interactions along a step of length s are generated from the artificial distribution (4.30) with first and second moments given by eqs. (4.28) and (4.29), which are determined by the transport mean free paths $\lambda_{comb,1}^{(s)}$ and $\lambda_{comb,2}^{(s)}$. To account (at least partially) for the energy dependence of these quantities we use a trick due to Alex Bielajew. The soft energy loss and angular deflection (which occur at the hinge) are considered as independent processes and are simulated in random order. That is, the soft angular deflection is evaluated for the energy at either the beginning or the end of the step, with equal probabilities. This is equivalent to assuming that the transport mean free paths $\lambda_{comb,1}^{(s)}(E)$ and $\lambda_{comb,2}^{(s)}(E)$ vary linearly with energy. The method is fairly accurate and computationally inexpensive provided only that the fractional energy loss along each step (which is of the order of C_2) is sufficiently small.

4.4 Generation of random tracks

Each simulated electron or positron history consists of a chronological succession of events. These can be either hard events, artificial soft events (hinges) or other relevant

stages of the particle history (such as its initial state, the crossing of an interface or the effective absorption after slowing down). The trajectory of the particle between a pair of successive events is straight and will be referred to as a "segment". We keep the term "step" to designate the portion of a track between two hard events, which consists of two segments and a hinge (when mixed simulation is effective).

Simulation with PENELOPE is controlled by the constants C_1 and C_2 [see eq. (4.85)] and the cutoff energies W_{cc} and W_{cr}. Hereafter, these four quantities will be referred to as simulation parameters. The parameter C_1, which determines the mean free path $\lambda_{el}^{(h)}$ between hard elastic events, should be small enough to ensure reliable simulation results. PENELOPE admits values of C_1 from 0 (detailed simulation) up to 0.2, which corresponds to a mean angular deflection $\langle\theta\rangle \sim 37$ deg after a steplength $\lambda_{el}^{(h)}$. The simulation parameter C_2 gives the maximum average fractional energy loss in a single step and it is effective only at high energies. From the discussion in section 4.3, it is clear that C_2 should also be small. PENELOPE allows values of C_2 between zero and 0.2. The cutoff energies W_{cc} and W_{cr} mainly influence the simulated energy distributions. The simulation speeds up by using larger cutoff energies, but if these are too large the simulated distributions may be somewhat distorted. In practice, simulated energy distributions are found to be quite insensitive to the adopted values of W_{cc} and W_{cr} when these are less than the bin width used to tally the energy distributions. Thus, the desired energy resolution determines the maximum allowed cutoff energies.

The combined effect of all soft elastic and stopping interactions in a step is simulated as a single artificial event or hinge, in which the particle changes its direction of movement and loses energy. When W_{cc} is less than the lowest oscillator resonance energy, the simulation of inelastic collisions becomes purely detailed, i.e. inelastic collisions do not contribute to the soft stopping power. On the other hand, the simulation of bremsstrahlung emission is only possible by means of a mixed scheme, because of the divergence of the DCS at $W = 0$ [see eq. (3.126)]. To test the accuracy of mixed algorithms, and also in studies of low-energy electron and positron transport (with, say, $E < 100$ keV), it may be convenient to perform strictly detailed simulations (see below). For this purpose, PENELOPE allows the user to switch off the emission of soft bremsstrahlung photons with energy less than 10 eV. This option is activated when the W_{cr} value selected by the user is negative, in which case the program sets $W_{cr} = 10$ eV, disregards soft bremsstrahlung events and simulates hard events (with $W > 10$ eV) in a detailed way. The generation of the angular deflection in artificial events is discontinued when the simulation of elastic and inelastic scattering becomes detailed (i.e. when $\lambda_{el}^{(h)} = \lambda_{el}$, $W_{cc} = 0$).

As indicated above, the length of the steps generated by PENELOPE is always less than s_{max}, an upper bound selected by the user. The simulation code limits the step length by placing delta interactions along the particle track. These are fictitious interactions that do not alter the state of the particle. Their only effect is to interrupt the sequence of simulation operations, which requires altering the values of inner control variables to permit resuming the simulation in a consistent way. The use of bounded step lengths is necessary to account for the energy dependence of the DCSs for soft interactions.

However, this is not the only reason for limiting the step length. Since energy losses and deflections at the hinges are sampled from artificial distributions, the number of hinges per primary track must be "statistically sufficient", i.e. larger than ~ 10, to smear off the unphysical details of the adopted artificial distributions. Therefore, when the particle is in a thin region, it is advisable to use a small value of $s_{\max}$ to make sure that the number of hinges within the material is sufficient. In PENELOPE, the parameter $s_{\max}$ can be varied freely during the course of the simulation of a single track. To ensure internal consistency, $s_{\max}$ is required to be less than $3\lambda_{\mathrm{T}}^{(\mathrm{h})}$. When the user-selected value is larger, the code sets $s_{\max} = 3\lambda_{\mathrm{T}}^{(\mathrm{h})}$; in this case, about 5 per cent of the sampled steps have lengths that exceed $s_{\max}$ and are terminated by a delta interaction. This slows down the simulation a little ($\sim$5%), but ensures that the energy dependence of $\lambda_{\mathrm{T}}^{(\mathrm{h})}$ is correctly accounted for. Instead of the $s_{\max}$ value set by the user, PENELOPE uses a random maximum step length [from a triangle distribution in the interval $(0,s_{\max})$] that averages to half the user's value; this is used to eliminate an artifact in the depth-dose distribution from parallel electron/positron beams near the entrance interface. Incidentally, limiting the step length is also necessary to perform simulation of electron/positron transport in external static electromagnetic fields (see appendix C).

The state of the particle immediately after an event is defined by its position coordinates $\mathbf{r}$, energy E and direction cosines of its direction of movement $\hat{\mathbf{d}}$, as seen from the laboratory reference frame. It is assumed that particles are locally absorbed when their energy becomes smaller than a preselected value E_{abs}; positrons are considered to annihilate after absorption. The practical generation of random electron and positron tracks in arbitrary material structures, which may consist of several homogeneous regions of different compositions separated by well-defined surfaces (interfaces), proceeds as follows:

(i) Set the initial position $\mathbf{r}$, kinetic energy E and direction of movement $\hat{\mathbf{d}}$ of the primary particle.

(ii) Determine the maximum allowed soft energy loss $\omega_{\max}$ along a step and set the value of inverse mean free path for hard events (see section 4.3). The results depend on the adopted $s_{\max}$, which can vary along the simulated track.

(iii) Sample the distance s to be travelled to the following hard event (or delta interaction) as

$$s = -\ln \xi / \Sigma_{\mathrm{h,max}}. \tag{4.119}$$

If $s > s_{\max}$, truncate the step by setting $s = s_{\max}$.

(iv) Generate the length $\tau = s\xi$ of the step to the next hinge. Let the particle advance this distance in the direction $\hat{\mathbf{d}}$: $\mathbf{r} \leftarrow \mathbf{r} + \tau\hat{\mathbf{d}}$.

(v) If the track has crossed an interface:
Stop it at the crossing point (i.e. redefine $\mathbf{r}$ as equal to the position of this point and set τ equal to the travelled distance).
Go to (ii) to continue the simulation in the new material, or go to (xi) if the new material is the outer vacuum.

(vi) Simulate the energy loss and deflection at the hinge. This step consists of two actions:

a) Sample the polar angular deflection $\mu = (1 - \cos\theta)/2$ from the distribution $F_a(s;\mu)$, eq. (4.30), corresponding to the current energy E. Sample the azimuthal scattering angle as $\phi = 2\pi\xi$. Perform a rotation $R(\theta,\phi)$ of the vector $\hat{\mathbf{d}}$ according to the sampled polar and azimuthal angular deflections (as described in section 1.4.2) to obtain the new direction: $\hat{\mathbf{d}} \leftarrow R(\theta,\phi)\hat{\mathbf{d}}$.

b) Sample the energy loss ω due to soft stopping interactions along the step s from the distribution $G_a(s;\omega)$, eqs. (4.59)-(4.63), and reduce the kinetic energy: $E \leftarrow E - \omega$.

These two actions are performed in random order to account for the energy dependence of the soft transport mean free paths (see section 4.3.3).

Go to (xi) if $E < E_{abs}$.

(vii) Let the particle advance the distance $s - \tau$ in the direction $\hat{\mathbf{d}}$: $\mathbf{r} \leftarrow \mathbf{r} + (s-\tau)\hat{\mathbf{d}}$.

(viii) Do as in (v).

(ix) If in step (iii) the step length was truncated, i.e. $s = s_{max}$, simulate a delta interaction.

Go to (ii).

(x) Simulate the hard event:

Sample the kind of interaction according to the point probabilities,

$$p_{el} = \frac{\mathcal{N}\sigma_{el}^{(h)}}{\Sigma_{h,max}}, \quad p_{in} = \frac{\mathcal{N}\sigma_{in}^{(h)}}{\Sigma_{h,max}}, \quad p_{br} = \frac{\mathcal{N}\sigma_{br}^{(h)}}{\Sigma_{h,max}}, \quad p_{\delta} = \frac{\mathcal{N}\sigma_{\delta}^{(h)}}{\Sigma_{h,max}},$$

$$\text{and} \quad p_{an} = \frac{\mathcal{N}\sigma_{an}^{(h)}}{\Sigma_{h,max}} \quad \text{in the case of positrons.} \tag{4.120}$$

If the event is a delta interaction, return to (ii).

Sample the polar scattering angle θ and the energy loss W from the corresponding DCS. Generate the azimuthal scattering angle as $\phi = 2\pi\xi$. Perform a rotation $R(\theta,\phi)$ of the vector $\hat{\mathbf{d}}$ to obtain the new direction: $\hat{\mathbf{d}} \leftarrow R(\theta,\phi)\hat{\mathbf{d}}$.

Reduce the kinetic energy of the particle: $E \leftarrow E - W$.

If, as a result of the interaction, a secondary particle is emitted in a direction $\hat{\mathbf{d}}_s$, with energy $E_s > E_{abs}$, store its initial state $(\mathbf{r}, E_s, \hat{\mathbf{d}}_s)$.

Go to (ii) if $E > E_{abs}$.

(xi) Simulate the tracks of the secondary electrons and photons produced by the primary particle (or by other secondaries previously followed) before starting a new primary track.

4.4.1 Stability of the simulation algorithm

The present simulation scheme for electrons/positrons is relatively stable under variations of the simulation parameters, due mostly to the effectiveness of the energy-loss corrections. This implies that the simulation parameters can be varied amply without practically altering the accuracy of the results. For the important case of low-energy electrons/positrons (with energies of the order of 500 keV or less), the relevant parameters are E_{abs}, C_1, W_{cc} and s_{max}, because C_2 is not effective (see fig. 4.3) and radiative

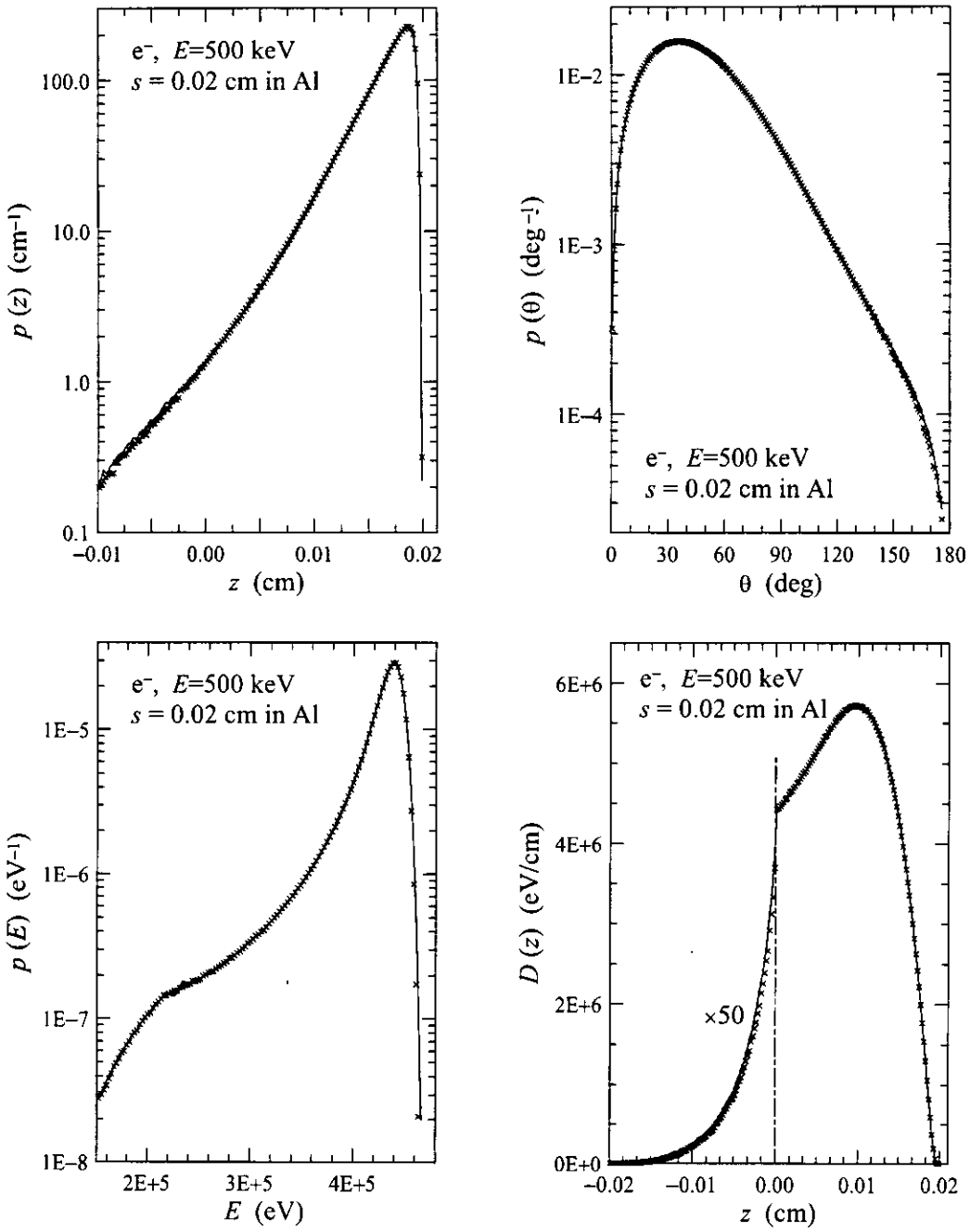


Figure 4.5: Results from the simulations of 500 keV electrons in aluminium described in the text. Crosses, detailed simulation; continuous curves, mixed simulation. $p(z)$ is the PDF of the z-coordinate of the final electron position, after travelling the prescribed 200 μm. $p(\theta)$ and $p(E)$ are the PDFs of the direction of motion (specified by the polar angle θ) and the kinetic energy E of the electrons at the end of the simulated tracks. The function $D(z)$ represents the "depth-dose" function, i.e. the average energy deposited in the material per unit length along the z-direction (the residual energy at the end of the track is not included in the dose).

emission is unimportant (hard bremsstrahlung events occur very seldom and, therefore, W_{cr} has no influence). The value of the parameter s_{max} is important to ensure the reliability of the results; a safe recipe is to set s_{max} equal to one tenth of the "expected track length" or less. Since the values of E_{abs} and W_{cc} are dictated by the characteristics of the considered experiment, it follows that the only "critical" parameter, with a direct influence on the speed of the simulation, is C_1. As mentioned above, PENELOPE accepts values of C_1 ranging from 0 (detailed simulation of elastic scattering) to 0.2.

In practice, the value of C_1 does not influence the *accuracy* of the simulation results when the other parameters are given "safe" values. This is illustrated in fig. 4.5, which displays results from simulations of 500 keV electrons in aluminium (infinite medium). Electrons started off from the origin of coordinates moving in the direction of the z axis. During the generation of each electron track, we scored the energy deposited at different "depths" (z-coordinate) to get the "depth-dose" distribution. The simulation of a track was discontinued when the electron had travelled a path length s equal to 200 μm, and the PDFs of the final electron energy and position coordinates were tallied. Notice that no secondary radiation was followed and that the kinetic energy of the electrons at $s = 200$ μm was not included in the dose distribution (i.e. the calculated "dose" does not represent the quantity that would be obtained from a measurement).

The results displayed in fig. 4.5 are from equivalent detailed and mixed simulations with $E_{abs} = 10$ keV and $s_{max} = 40$ μm. The detailed simulation was performed by setting $C_1 = C_2 = 0$, $W_{cc} = 0$ and $W_{cr} = -100$. Notice that when the user enters a negative value of the cutoff energy loss for radiative events, PENELOPE sets $W_{cr} = 10$ eV, disregards the emission of soft bremsstrahlung photons with $W < 10$ eV (which represents a negligible fraction of the stopping power) and simulates hard bremsstrahlung events as usually, i.e. in a detailed way. The mixed simulation results shown in fig. 4.5 were generated with $C_1 = C_2 = 0.2$, $W_{cc} = 1$ keV and $W_{cr} = -100$ (i.e. radiative events were described as in the detailed simulation).

In the detailed simulation, about 15 million electron tracks were generated by running a modified version of the code `PENSLAB.F` (see section 6.2.1) on a 666 MHz PII computer for 85 hours, which corresponds to a simulation speed of 49 tracks/s. The average numbers of elastic, inelastic and bremsstrahlung interactions that had to be simulated to produce each detailed track were 1297 and 1222 and 0.03, respectively. On the same computer, the mixed simulation generated 25 million tracks in about 2 hours, which represents a simulation speed of 3421 tracks/s, 71 times faster than that of detailed simulation. The reason is that, on average, there were only 2.4 hard elastic collisions, 6.2 hard inelastic collisions, 0.03 hard bremsstrahlung events and 6.8 delta interactions along each track. From fig. 4.5 we conclude that, in this case, the mixed algorithm is not only stable under variations of the parameter C_1 over the accepted range (0,0.2), but also provides results that are essentially equivalent to those from the detailed simulation. It is worth recalling that detailed simulation is nominally exact, the results are affected only by statistical uncertainties.

In general, our mixed simulation algorithm yields very accurate results (i.e. agree-

ing with those from detailed simulation) for electron and positron transport in infinite media, but not necessarily for limited geometries. The existence of interfaces poses considerable problems to condensed (class I) simulation, for which a satisfactory solution/approximation is not yet known. The present mixed (class II) algorithm handles interface crossing in a more accurate, but still approximate way. The rule to ensure accuracy for transport in the vicinity of interfaces is to use a small enough value of s_{max}.

Chapter 5

Constructive quadric geometry

Practical simulations of radiation transport in material systems involve two different kinds of operations, namely, physical (determination of the path length to the next interaction, random sampling of the different interactions) and geometrical (space displacements, interface crossings, ...). In the case of material systems with complex geometries, geometrical operations can take a large fraction of the simulation time. These operations are normally performed by dedicated subroutine packages, whose characteristics depend on the kind of algorithm used to simulate the interactions. The material system is assumed to consist of a number of homogeneous bodies limited by well-defined surfaces. The evolution of particles within each homogeneous body is dictated by the physical simulation routines, which operate as if particles were moving in an infinite medium with a given composition. Normally, the physical routines can handle a number of different media, whose interaction properties have been previously stored in memory. The job of the geometry routines is to steer the simulation of particle histories in the actual material system. They must determine the active medium, change it when the particle crosses an interface (i.e. a surface that separates two different media) and, for certain simulation algorithms, they must also keep control of the proximity of interfaces.

In this chapter we describe the FORTRAN subroutine package PENGEOM, which is adequate for detailed simulation algorithms (i.e. algorithms where all single interactions in the history of a particle are simulated in chronological succession). With these algorithms, the description of interface crossings is very simple: when the particle reaches an interface, its track is stopped just after entering a new material body and restarted again with the new active medium. This method (stopping and restarting a track when it crosses an interface) is applicable even when we have the same medium on both sides of the surface. That is, detailed simulations with a single homogeneous body and with the same body split into two parts by an arbitrary surface yield the same results (apart from statistical uncertainties).

As we have seen, detailed simulation is feasible only for photon transport and low-energy electron transport. For high-energy electrons and positrons, most Monte Carlo codes [e.g. ETRAN (Berger and Seltzer, 1988), ITS3 (Halbleib et al., 1992), EGS4 (Nel-

son et al., 1985), EGSnrc (Kawrakow and Rogers, 2000), GEANT (Brun et al., 1986)] have recourse to condensed (class I) or mixed (class II) simulation, where the global effect of multiple interactions along a path segment of a given length is evaluated using available multiple scattering theories. To avoid large step lengths that could place the particle inside a different medium, these condensed procedures require the evaluation of the distance from the current position to the nearest interface, an operation with a high computational cost (see e.g. Bielajew, 1995). The mixed procedure implemented in PENELOPE is, at least computationally, analogous to detailed simulation (it gives a "jump-and-knock" description of particle tracks). In fact, the structure of PENELOPE's tracking algorithm was designed to minimize the influence of the geometry on the transport physics. This algorithm operates independently of the proximity of interfaces and only requires knowledge of the material at the current position of the particle. As a consequence, the geometry package PENGEOM is directly linkable to PENELOPE. However, since PENGEOM does not evaluate the distance to the closest interface, it cannot be used with condensed simulation codes, such as those mentioned above.

Let us mention, in passing, that in simulations of high-energy photon transport complex geometries can be handled by means of relatively simple methods, which do not require control of interface crossings (see e.g. Snyder et al., 1969). Unfortunately, similar techniques are not applicable to electron and positron transport, mainly because these particles have much shorter track lengths and, hence, the transport process is strongly influenced by inhomogeneities of the medium. With the analogue simulation scheme adopted in PENELOPE, it is necessary to determine when a particle track crosses an interface, not only for electrons and positrons but also for photons.

PENGEOM evolved from a subroutine package of the same name provided with the 1996.02.29 version of the PENELOPE code system. This package was aimed at describing simple structures with a small number of homogeneous bodies limited by quadric surfaces. Although it was robust and very flexible, its speed deteriorated rapidly when the number of surfaces increased. The need for developing a more efficient geometry package became evident when we started to use PENELOPE to simulate radiation transport in accelerator heads (the description of which requires of the order of 100 surfaces) or in studies of total body irradiation (the definition of a realistic anthropomorphic phantom may involve a few hundred surfaces).

With PENGEOM we can describe any material system consisting of homogeneous bodies limited by quadric surfaces. To speed up the geometry operations, the bodies of the material system can be grouped into modules (connected volumes, limited by quadric surfaces, that contain one or several bodies); modules can in turn form part of larger modules, and so on. This hierarchic modular structure allows a reduction of the work of the geometry routines, which becomes more effective when the complexity of the system increases.

Except for trivial cases, the correctness of the geometry definition is difficult to check and, moreover, 3D structures with interpenetrating bodies are difficult to visualize. A pair of programs, named GVIEW2D and GVIEW3D, have been written to display the ge-

ometry on the computer screen. These programs use specific computer graphics software and, therefore, they are not portable. The executable files included in the PENELOPE distribution package run on IBM-compatible personal computers under Microsoft Windows 9x; they are simple and effective tools for debugging the geometry definition file.

5.1 Rotations and translations

The definition of various parts of the material system (quadric surfaces in reduced form and modules) involves rotations and translations. To describe these transformations, we shall adopt the active point of view: the reference frame remains fixed and only the space points (vectors) are translated or rotated.

In what follows, and in the computer programs, all lengths are in cm. The position and direction of movement of a particle are referred to the laboratory coordinate system, a Cartesian reference frame which is defined by the position of its origin of coordinates and the unit vectors $\hat{\mathbf{x}} = (1,0,0)$, $\hat{\mathbf{y}} = (0,1,0)$ and $\hat{\mathbf{z}} = (0,0,1)$ along the directions of its axes.

A translation $\mathcal{T}(\mathbf{t})$, defined by the displacement vector $\mathbf{t} = (t_x, t_y, t_z)$, transforms the vector $\mathbf{r} = (x, y, z)$ into

$$\mathcal{T}(\mathbf{t})\,\mathbf{r} = \mathbf{r} + \mathbf{t} = (x + t_x, y + t_y, z + t_z). \tag{5.1}$$

Evidently, the inverse translation $\mathcal{T}^{-1}(\mathbf{t})$ corresponds to the displacement vector $-\mathbf{t}$, i.e. $\mathcal{T}^{-1}(\mathbf{t}) = \mathcal{T}(-\mathbf{t})$.

A rotation $\mathcal{R}$ is defined through the Euler angles ω, θ and ϕ, which specify a sequence of rotations about the coordinate axes[1]: first a rotation of angle ω about the z-axis, followed by a rotation of angle θ about the y-axis and, finally, a rotation of angle ϕ about the z-axis. A positive rotation about a given axis would carry a right-handed screw in the positive direction along that axis. Positive (negative) angles define positive (negative) rotations.

The rotation $\mathcal{R}(\omega, \theta, \phi)$ transforms the vector $\mathbf{r} = (x, y, z)$ into a vector

$$\mathbf{r}' = \mathcal{R}(\omega, \theta, \phi)\,\mathbf{r} = (x', y', z'), \tag{5.2}$$

whose coordinates are given by

$$\begin{pmatrix} x' \\ y' \\ z' \end{pmatrix} = R(\omega, \theta, \phi) \begin{pmatrix} x \\ y \\ z \end{pmatrix}, \tag{5.3}$$

[1]This definition of the Euler angles is the one usually adopted in Quantum Mechanics (see e.g. Edmonds, 1960).

where

$$R(\omega,\theta,\phi) = \begin{pmatrix} R_{xx} & R_{xy} & R_{xz} \\ R_{yx} & R_{yy} & R_{yz} \\ R_{zx} & R_{zy} & R_{zz} \end{pmatrix} \tag{5.4}$$

is the rotation matrix. To obtain its explicit form, we recall that the matrices for rotations about the z- and y-axes are

$$R_z(\phi) = \begin{pmatrix} \cos\phi & -\sin\phi & 0 \\ \sin\phi & \cos\phi & 0 \\ 0 & 0 & 1 \end{pmatrix} \quad \text{and} \quad R_y(\theta) = \begin{pmatrix} \cos\theta & 0 & \sin\theta \\ 0 & 1 & 0 \\ -\sin\theta & 0 & \cos\theta \end{pmatrix}, \tag{5.5}$$

respectively. Hence,

$$\begin{aligned} R(\omega,\theta,\phi) &= R_z(\phi)R_y(\theta)R_z(\omega) \\ &= \begin{pmatrix} \cos\phi & -\sin\phi & 0 \\ \sin\phi & \cos\phi & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} \cos\theta & 0 & \sin\theta \\ 0 & 1 & 0 \\ -\sin\theta & 0 & \cos\theta \end{pmatrix} \begin{pmatrix} \cos\omega & -\sin\omega & 0 \\ \sin\omega & \cos\omega & 0 \\ 0 & 0 & 1 \end{pmatrix} \end{aligned}$$

$$= \begin{pmatrix} \cos\phi\cos\theta\cos\omega - \sin\phi\sin\omega & -\cos\phi\cos\theta\sin\omega - \sin\phi\cos\omega & \cos\phi\sin\theta \\ \sin\phi\cos\theta\cos\omega + \cos\phi\sin\omega & -\sin\phi\cos\theta\sin\omega + \cos\phi\cos\omega & \sin\phi\sin\theta \\ -\sin\theta\cos\omega & \sin\theta\sin\omega & \cos\theta \end{pmatrix}. \tag{5.6}$$

The inverse of the rotation $\mathcal{R}(\omega,\theta,\phi)$ is $\mathcal{R}(-\phi,-\theta,-\omega)$ and its matrix is the transpose of $R(\omega,\theta,\phi)$, i.e.

$$R^{-1}(\omega,\theta,\phi) = R(-\phi,-\theta,-\omega) = R_z(-\omega)R_y(-\theta)R_z(-\phi) = R^{\mathrm{T}}(\omega,\theta,\phi). \tag{5.7}$$

Let us now consider transformations $\mathcal{C} = \mathcal{T}(\mathbf{t})\,\mathcal{R}(\omega,\theta,\phi)$ that are products of a rotation $\mathcal{R}(\omega,\theta,\phi)$ and a translation $\mathcal{T}(\mathbf{t})$. $\mathcal{C}$ transforms a point $\mathbf{r}$ into

$$\mathbf{r}' = \mathcal{C}(\mathbf{r}) = \mathcal{T}(\mathbf{t})\,\mathcal{R}(\omega,\theta,\phi)\,\mathbf{r} \tag{5.8}$$

or, in matrix form,

$$\begin{pmatrix} x' \\ y' \\ z' \end{pmatrix} = R(\omega,\theta,\phi) \begin{pmatrix} x \\ y \\ z \end{pmatrix} + \begin{pmatrix} t_x \\ t_y \\ t_z \end{pmatrix}. \tag{5.9}$$

Notice that the order of the factors does matter; the product of the same factors in reverse order $\mathcal{D} = \mathcal{R}(\omega,\theta,\phi)\,\mathcal{T}(\mathbf{t})$ transforms $\mathbf{r}$ into a point $\mathbf{r}' = \mathcal{D}(\mathbf{r})$ with coordinates

$$\begin{pmatrix} x' \\ y' \\ z' \end{pmatrix} = R(\omega,\theta,\phi) \begin{pmatrix} x + t_x \\ y + t_y \\ z + t_z \end{pmatrix}. \tag{5.10}$$

Given a function $F(\mathbf{r})$, the equation $F(\mathbf{r}) = 0$ defines a surface in implicit form. We can generate a new surface by applying a rotation $\mathcal{R}(\omega, \theta, \phi)$ followed by a translation $\mathcal{T}(\mathbf{t})$ (we shall always adopt this order). The implicit equation of the transformed surface is

$$G(\mathbf{r}) = F\left[\mathcal{R}^{-1}(\omega, \theta, \phi)\,\mathcal{T}^{-1}(\mathbf{t})\,\mathbf{r}\right] = 0, \tag{5.11}$$

which simply expresses the fact that $G(\mathbf{r})$ equals the value of the original function at the point $\mathbf{r}' = \mathcal{R}^{-1}(\omega, \theta, \phi)\,\mathcal{T}^{-1}(\mathbf{t})\,\mathbf{r}$ that transforms into $\mathbf{r}$.

5.2 Quadric surfaces

As already mentioned, the material system consists of a number of homogeneous bodies, defined by their composition (material) and limiting surfaces. For practical reasons, all limiting surfaces are assumed to be quadrics given by the implicit equation

$$\begin{aligned} F(x,y,z) = {} & A_{xx}x^2 + A_{xy}xy + A_{xz}xz + A_{yy}y^2 + A_{yz}yz + A_{zz}z^2 \\ & + A_x x + A_y y + A_z z + A_0 = 0, \end{aligned} \tag{5.12}$$

which includes planes, pairs of planes, spheres, cylinders, cones, ellipsoids, paraboloids, hyperboloids, etc. In practice, limiting surfaces are frequently known in "graphical" form and it may be very difficult to obtain the corresponding quadric parameters. Try with a simple example: calculate the parameters of a circular cylinder of radius R such that its symmetry axis goes through the origin and is parallel to the vector (1,1,1). To facilitate the definition of the geometry, each quadric surface can be specified either through its implicit equation or by means of its reduced form, which defines the "shape" of the surface (see fig. 5.1), and a few simple geometrical transformations.

A reduced quadric is defined by the expression

$$F_r(x,y,z) = I_1x^2 + I_2y^2 + I_3z^2 + I_4z + I_5 = 0, \tag{5.13}$$

where the coefficients (indices) I_1 to I_5 can only take the values -1, 0 or 1. Notice that reduced quadrics have central symmetry about the z-axis, i.e. $F_r(-x,-y,z) = F_r(x,y,z)$. The possible (real) reduced quadrics are given in table 5.1.

A general quadric is obtained from the corresponding reduced form by applying the following transformations (in the quoted order)[2].

(i) An expansion along the directions of the axes, defined by the scaling factors `X-SCALE`$= a$, `Y-SCALE`$= b$ and `Z-SCALE`$= c$. The equation of the scaled quadric is

$$F_s(x,y,z) = I_1\left(\frac{x}{a}\right)^2 + I_2\left(\frac{y}{b}\right)^2 + I_3\left(\frac{z}{c}\right)^2 + I_4\frac{z}{c} + I_5 = 0. \tag{5.14}$$

[2]Keywords used to denote the various parameters in the geometry definition file are written in typewriter font, e.g. `X-SCALE`). See section 5.4.

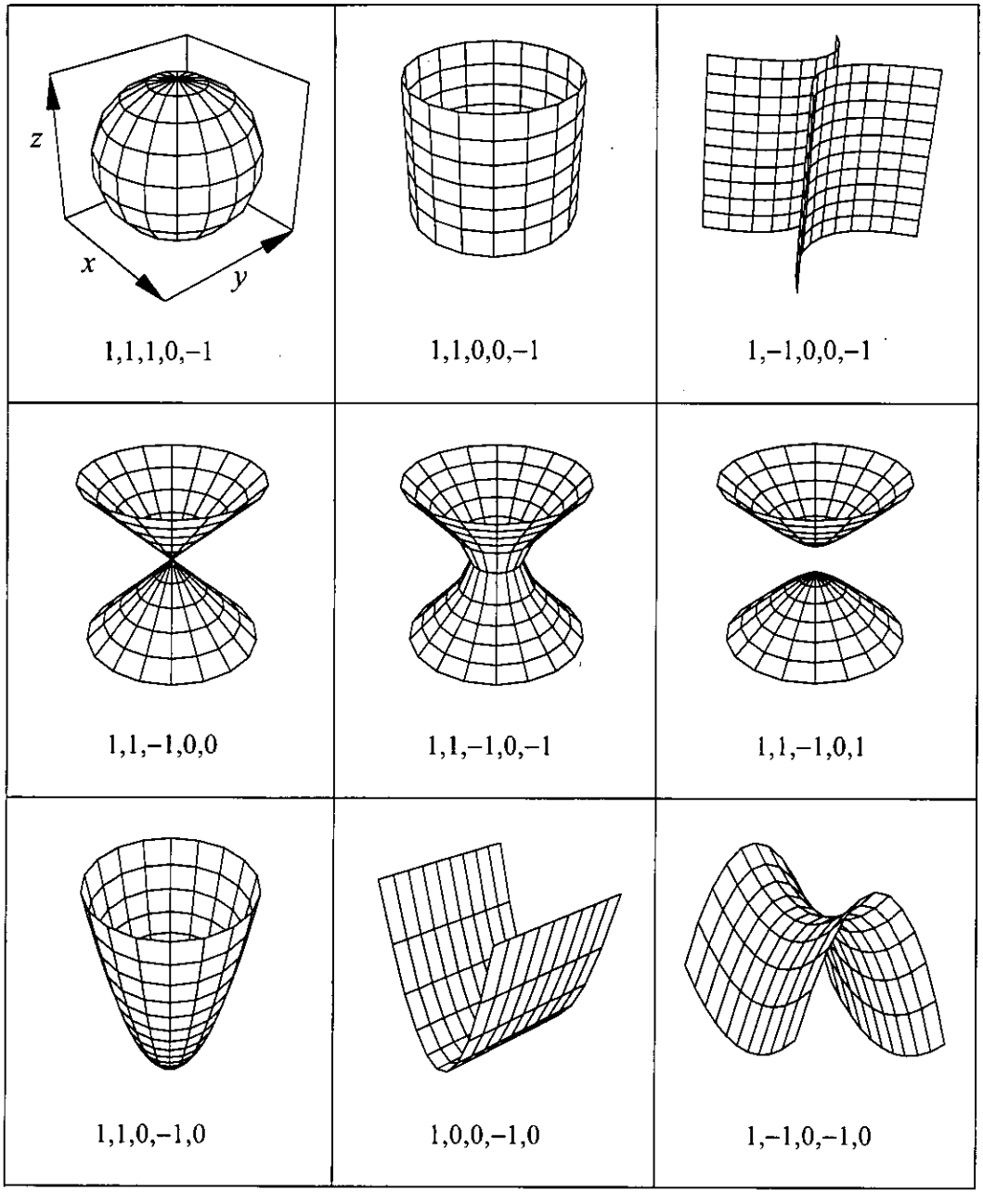


Figure 5.1: Non-planar reduced quadric surfaces and their indices [see eq. (5.13)]. In all cases, the perspective is the same as for the sphere.

Table 5.1: Reduced quadrics.

Reduced form	Indices					Quadric
$z-1=0$	0	0	0	1	−1	plane
$z^2-1=0$	0	0	1	0	−1	pair of parallel planes
$x^2+y^2+z^2-1=0$	1	1	1	0	−1	sphere
$x^2+y^2-1=0$	1	1	0	0	−1	cylinder
$x^2-y^2-1=0$	1	−1	0	0	−1	hyperbolic cylinder
$x^2+y^2-z^2=0$	1	1	−1	0	0	cone
$x^2+y^2-z^2-1=0$	1	1	−1	0	−1	one sheet hyperboloid
$x^2+y^2-z^2+1=0$	1	1	−1	0	1	two sheet hyperboloid
$x^2+y^2-z=0$	1	1	0	−1	0	paraboloid
$x^2-z=0$	1	0	0	−1	0	parabolic cylinder
$x^2-y^2-z=0$	1	−1	0	−1	0	hyperbolic paraboloid

... and permutations of x, y and z that preserve the central symmetry with respect to the z-axis.

For instance, this transforms the reduced sphere into an ellipsoid with semiaxes equal to the scaling factors.

(ii) A rotation, $\mathcal{R}(\omega,\theta,\phi)$, defined through the Euler angles `OMEGA`$=\omega$, `THETA`$=\theta$ and `PHI`$=\phi$. Notice that the rotation $\mathcal{R}(\omega,\theta,\phi)$ transforms a plane perpendicular to the z-axis into a plane perpendicular to the direction with polar and azimuthal angles `THETA` and `PHI`, respectively. The first Euler angle, ω has no effect when the initial (scaled) quadric is symmetric about the z-axis.

(iii) A translation, defined by the components of the displacement vector $\mathbf{t}$ (`X-SHIFT`$=t_x$, `Y-SHIFT`$=t_y$, `Z-SHIFT`$=t_z$).

A quadric is completely specified by giving the set of indices (I_1, I_2, I_3, I_4, I_5), the scale factors (`X-SCALE`, `Y-SCALE`, `Z-SCALE`), the Euler angles (`OMEGA`, `THETA`, `PHI`) and the displacement vector (`X-SHIFT`, `Y-SHIFT`, `Z-SHIFT`). Any quadric surface can be expressed in this way. The implicit equation of the quadric is obtained as follows. We define the matrix

$$\mathcal{A}=\begin{pmatrix} A_{xx} & \frac{1}{2}A_{xy} & \frac{1}{2}A_{xz} \\ \frac{1}{2}A_{xy} & A_{yy} & \frac{1}{2}A_{yz} \\ \frac{1}{2}A_{xz} & \frac{1}{2}A_{yz} & A_{zz} \end{pmatrix} \tag{5.15}$$

and write the generic quadric equation (5.12) in matrix form

$$\mathbf{r}^{\mathrm{T}}\mathcal{A}\,\mathbf{r}+\mathbf{A}^{\mathrm{T}}\mathbf{r}+A_0=0, \tag{5.16}$$

where $\mathbf{r}$ and $\mathbf{A} \equiv (A_x, A_y, A_z)$ are considered here as one-column matrices. Notice that the matrix $\mathcal{A}$ is symmetric ($\mathcal{A}^T = \mathcal{A}$). Expressing the scaled quadric (5.14) in the form (5.16), the equation for the rotated and shifted quadric is [see eq. (5.11)]

$$(\mathbf{r} - \mathbf{t})^T R\mathcal{A}R^T (\mathbf{r} - \mathbf{t}) + (R\mathbf{A})^T(\mathbf{r} - \mathbf{t}) + A_0 = 0, \tag{5.17}$$

which can be written in the generic form (5.16)

$$\mathbf{r}^T\mathcal{A}'\mathbf{r} + \mathbf{A}'^T\mathbf{r} + A_0' = 0 \tag{5.18}$$

with

$$\mathcal{A}' = R\mathcal{A}R^T, \qquad \mathbf{A}' = R\mathbf{A} - 2\mathcal{A}'\mathbf{t}, \qquad A_0' = A_0 + \mathbf{t}^T(\mathcal{A}'\mathbf{t} - R\mathbf{A}). \tag{5.19}$$

From these relations, the parameters of the implicit equation (5.12) are easily obtained.

A quadric surface $F(x, y, z) = 0$ divides the space into two exclusive regions that are identified by the sign of $F(x, y, z)$, the surface side pointer. A point with coordinates (x_0, y_0, z_0) is said to be inside the surface if $F(x_0, y_0, z_0) \leq 0$ (side pointer $= -1$), and outside it if $F(x_0, y_0, z_0) > 0$ (side pointer $= +1$).

5.3 Constructive quadric geometry

A body is defined as a space volume limited by quadric surfaces and filled with a homogeneous material. To specify a body we have to define its limiting quadric surfaces $F(\mathbf{r}) = 0$, with corresponding side pointers ($+1$ or -1), and its composition (i.e. the integer label used by PENELOPE to identify the material). It is considered that bodies are defined in "ascending", exclusive order so that previously defined bodies effectively delimit the new ones. This is convenient e.g. to describe bodies with inclusions. The work of the geometry routines is much easier when bodies are completely defined by their limiting surfaces, but this is not always possible or convenient for the user. The example in section 5.7 describes an arrow inside a sphere (fig. 5.2); the arrow is defined first so that it limits the volume filled by the material inside the sphere. It is impossible to define the hollow sphere (as a single body) by means of only its limiting quadric surfaces. It is clear that, by defining a conveniently large number of surfaces and bodies, we can describe any quadric geometry.

The subroutine package PENGEOM contains a subroutine, named `LOCATE`, that "locates" a point $\mathbf{r}$, i.e. determines the body that contains it, if any. The obvious method is to compute the side pointers [i.e. the sign of $F(\mathbf{r})$] for *all* surfaces and, then, explore the bodies in ascending order looking for the first one that fits the given side pointers. This brute force procedure was used in older versions of PENGEOM; it has the advantage of being robust (and easy to program) but becomes too slow for complex systems. A second subroutine, named `STEP`, "moves" the particle from a given position $\mathbf{r}_0$ within a body B a certain distance s in a given direction $\hat{\mathbf{d}}$. `STEP` also checks if the particle leaves the active medium and, when this occurs, stops the particle just after entering

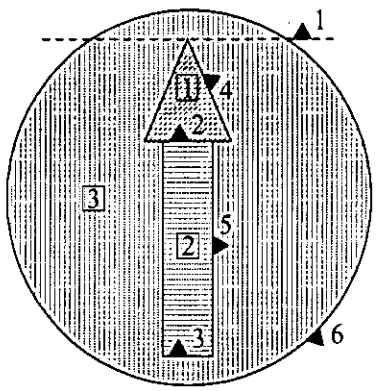


Figure 5.2: Example of simple quadric geometry; an arrow within a sphere (the corresponding definition file is given in section 5.7). The solid triangles indicate the outside of the surfaces (side pointer = +1). Numbers in squares indicate bodies.

the new material. To do this, we must determine the intersections of the track segment $\mathbf{r}_0 + t\hat{\mathbf{d}}$ $(0 < t \leq s)$ with *all* the surfaces that limit the body B (including those that limit other bodies that limit B), and check if the final position $\mathbf{r}_0 + s\hat{\mathbf{d}}$ remains in B or not. The reason for using only quadric surfaces is that these intersections are easily calculated by solving a quadratic equation.

Notice that bodies can be concave, i.e., the straight segment joining any two points in a body may not be wholly contained in the body. Hence, even when the final position of the particle lies within the initial body, we must analyze all the intersections of the path segment with the limiting surfaces of B and check if the particle has left the body after any of the intersections. When the particle leaves the initial body, say after travelling a distance s' $(< s)$, we have to locate the point $\mathbf{r}' = \mathbf{r}_0 + s'\hat{\mathbf{d}}$. The easiest method consists of computing the side pointers of *all* surfaces of the system at $\mathbf{r}'$, and determining the body B' that contains $\mathbf{r}'$ by analyzing the side pointers of the different bodies in ascending order. It is clear that, for complex geometries, this is a very slow process. We can speed it up by simply disregarding those elements of the system that cannot be reached in a single step (e.g. bodies that are "screened" by other bodies). Unfortunately, as a body can be limited by all the other bodies that have been defined previously, the algorithm can be improved only at the expense of providing it with additional information. We shall adopt a simple strategy that consists of lumping groups of bodies together to form modules.

A module is defined as a connected volume[3], limited by quadric surfaces, that contains one or several bodies. A module can contain other modules, which will be referred to as submodules of the first. The volume of a module is filled with a homogeneous medium, which automatically fills the cavities of the module (i.e. volumes that do not

[3]A space volume is said to be connected when any two points in the volume can be joined by an arc of curve that is completely contained within the volume.

correspond to a body or to a submodule); these filled cavities are considered as a single new body. A body that is connected and limited only by surfaces can be declared either as a body or as a module. For the sake of simplicity, modules are required to satisfy the following conditions: 1) the bodies and submodules of a module must be completely contained within the parent module (i.e. it is not allowed to have portions of bodies or submodules that lie outside the module) and 2) a submodule of a module cannot overlap with other submodules and bodies of the same module (this is necessary to make sure that a particle can only enter or leave a module through its limiting surfaces). Notice however, that the bodies of a module are still assumed to be defined in ascending order, i.e. a body is limited by its surfaces and by the previously defined bodies *of the same module*, so that inclusions and interpenetrating bodies can be easily defined. Of course, overlapping bodies must be in the same module.

A module (with its possible submodules) can represent a rigid part (e.g. a radioactive source, an accelerator head, a detector, a phantom, etc.) of a more complex material system. To facilitate the definition of the geometry, it is useful to allow free translations and rotations of the individual modules. The definition of a module (see below) includes the parameters of a rotation $\mathcal{R}(\omega, \theta, \phi)$ and a translation $\mathcal{T}(\mathrm{t})$, which are optional and serve to modify the position and orientation of the module (and its submodules) with respect to the laboratory reference frame. As before, the rotation is applied first. All submodules and bodies of the same module are shifted and rotated together.

In practical simulations, it may be useful to limit the region of space where particles have to be transported. For instance, to simulate the response of a detector with a given photon source, it is advisable to stop the simulation of a particle when it is far enough from the detector. This can be done automatically by considering an "enclosure" of the material system, which is defined as a module that contains the complete system. If such a covering module is not explicitly defined, the subroutines set the enclosure as a sphere of 10^{15} cm radius. It is assumed that there is perfect vacuum outside the enclosure, and in any inner volume that is not a part of a body or of a filled module. Hence, particles that leave the enclosure are lost and will never return to the material system.

For programming purposes, it is useful to imagine each module as the mother of its bodies and submodules, and as the daughter of the module that contains it. We thus have a kind of genealogical tree with various generations of modules and bodies (see fig. 5.3). The first generation reduces to the enclosure (which is the only motherless module). The members of the second generation are bodies and modules that are daughters of the enclosure. The n-th generation consists of modules and bodies whose mothers belong to the $(n-1)$-th generation. Each module is defined by its limiting surfaces (which determine the border with the external world) and those of their descendants (which determine the module's internal structure); this is not true for bodies (childless members of the tree), which can be limited either by surfaces, by other sister bodies or by a combination of both. A body that is limited only by surfaces can be defined as a module, which has the advantage of allowing free rotation and translation.

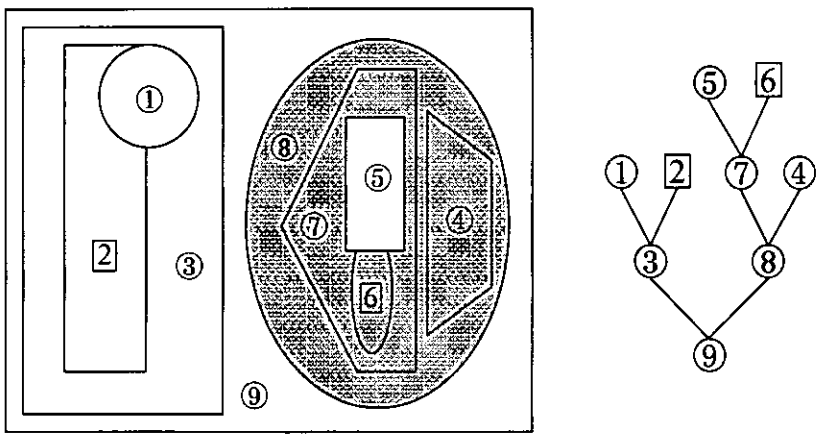


Figure 5.3: Planar cut of a geometry example, showing modules (number labels in circles) and bodies (number labels in squares), and the associated genealogical tree. Notice that a module can always be defined as a body limited by their submodules and daughter bodies, but this affects the structure of the genealogical tree and, therefore, the efficiency (speed) of the geometry operations.

5.4 Geometry definition file

The geometry is defined from the input file (`UNIT=IRD`). In principle, this permits the simulation of different geometries by using the same main program. The input file consists of a series of data sets, which define the different elements (surfaces, bodies and modules). A data set consists of a number of strictly formatted text lines; it starts and ends with a separation line filled with zeros. The first line after each separation line must start with one of the defining 8- character strings "`SURFACE-`", "`BODY----`", "`MODULE--`", "`END-----`" or "`INCLUDE-`" (here, blank characters are denoted by "-"; they are essential!). Informative text (as many lines as desired) can be written at the beginning of the file, before the first separation line. A line starting with the string "`END-----`" after a separation line discontinues the reading of geometry data. Each element is identified by its type (surface, body or module) and a three-digit integer label. Although the element label can be given an arbitrary value (-99 to 999) in the input file, PENGEOM redefines it so that elements of a given kind are numbered consecutively, according to their input order. Notice that bodies and modules are considered as elements of the same kind (i.e. assigning the same label to a body and to a module will cause an error of the reading routine).

In the input file, numerical quantities must be written within the parentheses in the specified format. All lengths are in cm; angles can be given in either degrees (`DEG`) or radians (`RAD`). When angles are in degrees, it is not necessary to specify the unit. The parameters in each data set can be entered in any order. They can even be defined several times, in which case, only the last input value is accepted. This is useful, e.g.

to study variations caused by changing these parameters without duplicating the input file. Comments can be written at the end of each line, at the right of the last keyword or after the closing parenthesis of numerical fields.

- The format of the data set of a surface defined in reduced form is the following,

```
0000000000000000000000000000000000000000000000000000000000000000000000
SURFACE ( I3)   TEXT DESCRIBING THE SURFACE ...
INDICES=(I2,I2,I2,I2,I2)
X-SCALE=(        E22.15        , I3)              (DEFAULT=1.0)
Y-SCALE=(        E22.15        , I3)              (DEFAULT=1.0)
Z-SCALE=(        E22.15        , I3)              (DEFAULT=1.0)
  OMEGA=(        E22.15        , I3) DEG          (DEFAULT=0.0)
  THETA=(        E22.15        , I3) DEG          (DEFAULT=0.0)
    PHI=(        E22.15        , I3) RAD          (DEFAULT=0.0)
X-SHIFT=(        E22.15        , I3)              (DEFAULT=0.0)
Y-SHIFT=(        E22.15        , I3)              (DEFAULT=0.0)
Z-SHIFT=(        E22.15        , I3)              (DEFAULT=0.0)
0000000000000000000000000000000000000000000000000000000000000000000000
```

  o Surface parameters are optional and can be entered in any order. Default values are assigned to parameters not defined in the input file. Thus, to define an elliptic cylinder centred on the z-axis, only the parameters `X-SCALE` and `Y-SCALE` are required. Notice that scale parameters must be greater than zero.
  o The `I3` value following each parameter must be set equal to zero (or negative) to make the parameter value effective. When this field contains a positive integer `IP`, the parameter is set equal to the value stored in the `IP`-th component of the array `PARINP`, an input argument of subroutine `GEOMIN` (see section 5.6). This permits the user to modify the geometry parameters from the `MAIN` program.

- Limiting surfaces can also be defined in implicit form. When a quadric surface is defined in this way, the indices must be set to zero; this switches the reading subroutine `GEOMIN` to implicit mode. The format of an implicit surface data set is

```
0000000000000000000000000000000000000000000000000000000000000000000000
SURFACE ( I3)   TEXT DESCRIBING THE SURFACE ...
INDICES=( 0, 0, 0, 0, 0)
    AXX=(        E22.15        , I3)              (DEFAULT=0.0)
    AXY=(        E22.15        , I3)              (DEFAULT=0.0)
    AXZ=(        E22.15        , I3)              (DEFAULT=0.0)
    AYY=(        E22.15        , I3)              (DEFAULT=0.0)
    AYZ=(        E22.15        , I3)              (DEFAULT=0.0)
    AZZ=(        E22.15        , I3)              (DEFAULT=0.0)
     AX=(        E22.15        , I3)              (DEFAULT=0.0)
     AY=(        E22.15        , I3)              (DEFAULT=0.0)
     AZ=(        E22.15        , I3)              (DEFAULT=0.0)
     A0=(        E22.15        , I3)              (DEFAULT=0.0)
0000000000000000000000000000000000000000000000000000000000000000000000
```

  o Surface parameters are optional and can be entered in any order. The default value 0.0 is assigned to parameters not defined in the input file.
  o The `I3` value following each parameter must be negative or zero to make the parameter value effective. Otherwise, its actual value will be set by subroutine `GEOMIN`.

- The format of a body data set is

```
0000000000000000000000000000000000000000000000000000000000000000000000
BODY     ( I3)   TEXT DESCRIBING THE BODY ...
MATERIAL( I3)
SURFACE ( I3), SIDE POINTER=(I2)
SURFACE ( I3), SIDE POINTER=(I2) ...
BODY     ( I3)
BODY     ( I3) ...
0000000000000000000000000000000000000000000000000000000000000000000000
```

  o The indicator of each material (2nd line) must agree with the convention adopted in PENELOPE. Void inner volumes can be described as material bodies with MATERIAL set to 0.

  o A line is required to define each limiting surface, with its side pointer, and each limiting body. Limiting surfaces and bodies can be entered in any order.

  o Bodies are assumed to be defined in ascending order so that, in principle, it would not be necessary to declare the limiting bodies. However, to speed up the calculations, it is required to declare explicitly all the elements (surfaces and bodies) that actually limit the body that is being defined. Omission of a limiting body will cause inconsistencies unless the materials in the limiting and the limited bodies are the same.

- The format for the definition of a module is the following:

```
0000000000000000000000000000000000000000000000000000000000000000000000
MODULE   ( I3)   TEXT DESCRIBING THE MODULE...
MATERIAL( I3)
SURFACE ( I3), SIDE POINTER=(I2)
SURFACE ( I3), SIDE POINTER=(I2) ...
BODY     ( I3)
BODY     ( I3) ...
MODULE   ( I3)
MODULE   ( I3) ...
1111111111111111111111111111111111111111111111111111111111111111111111
  OMEGA=(        E22.15        , I3) DEG           (DEFAULT=0.0)
  THETA=(        E22.15        , I3) DEG           (DEFAULT=0.0)
    PHI=(        E22.15        , I3) RAD           (DEFAULT=0.0)
X-SHIFT=(        E22.15        , I3)               (DEFAULT=0.0)
Y-SHIFT=(        E22.15        , I3)               (DEFAULT=0.0)
Z-SHIFT=(        E22.15        , I3)               (DEFAULT=0.0)
0000000000000000000000000000000000000000000000000000000000000000000000
```

  o The material (which must be explicitly declared) fills the cavities of the module. As in the case of bodies, MATERIAL = 0 corresponds to vacuum.

  o The limiting surfaces must define a connected volume. All inner bodies and modules (submodules) must be declared. Notice that these cannot extend outside the module's volume and that a submodule cannot overlap with the other submodules and bodies.

  o Limiting surfaces, inner bodies and submodules can be entered in any order.

  o The enclosure of the system can be defined as a module (the last in the geometry file). The enclosure must contain all modules and all bodies that do not belong to modules. Otherwise, the subroutines will define a new enclosure (a sphere with

10^{15} cm radius) that is supposed to satisfy this condition. Make sure that the system does not contain bodies that extend outside the enclosure (this normally will cause errors).

o The rotation and the translation are optional and apply to all elements of the module. The line filled with 1's ends the definition of elements and starts that of transformation parameters (it can be skipped if no transformation parameters are entered). The I3 value following each parameter has the same meaning as above. It has to be negative or zero to make the parameter value effective; otherwise, the parameter must be set from the main program.

A single surface can be used to define several bodies and/or submodules in the same module; unnecessary duplication of a surface reduces the calculation speed. Notice, however, that rotation or translation of a module modifies all the surfaces of its descendants and, therefore, a transformed surface must be redefined to be used again. Thus, if the system contains two identical modules in different positions (e.g. two detectors in a coincidence experiment), each of them must be defined explicitly. This does not require too much editing work; after generating the first of the two modules, we can just duplicate its definition data sets and change their labels.

The INCLUDE option allows inserting a predefined structure (e.g. a scintillation detector, an encapsulated nuclear source, ...) within the geometry file. The inserted structure is defined by a complete definition file (i.e. ending with an "END-----" line). The labels of the objects in the included file must be different from the labels used in the main file and in any other included file. The format of an INCLUDE block is the following,

```
0000000000000000000000000000000000000000000000000000000000000000
INCLUDE
FILE=   (filename.ext)
0000000000000000000000000000000000000000000000000000000000000000
```

The name of the included file must be written between the parentheses. It may be up to twelve characters long; if it is shorter, the blanks must be left at the right end of the field. Only one-level INCLUDES are allowed, i.e. an included file cannot contain any INCLUDE blocks.

The definition of the geometry may seem somewhat more laborious than with combinatorial methods, where the system is described by combining basic bodies of several simple shapes [see Jenkins et al. (1988) and references therein]. In practice, however, defining the various surfaces that limit a body may be more convenient, and intuitive, than considering all the parameters needed to specify that body. The example of a right elliptical cylinder, which needs 9 parameters, is quite illustrative. With our method, this body can be defined as a module by means of two planes perpendicular to the z-axis (only one parameter if the base is the $z = 0$ plane) and a scaled cylinder (2 parameters); the rotation (3 parameters) of the module gives the required orientation and the translation (3 parameters) puts it in the required position. The definition as a proper body requires defining the three surfaces that limit the cylinder in its actual position,

which is a bit more inconvenient. In any case, the important issue is not how to define the geometry, but the amount of computation needed to follow a particle through the material system.

5.5 The subroutine package PENGEOM

The package PENGEOM consists of the following subroutines;

- `SUBROUTINE GEOMIN(PARINP,NPINP,NMAT,NBOD,IRD,IWR)`
  Reads geometry data from the input file and initializes the geometry package.

  o Input arguments:
  `PARINP`: Array containing optional parameters, which may replace the ones entered from the input file. This array must be declared in the `MAIN` program, even when `NPINP` = 0.
  `NPINP`: Number of parameters defined in `PARINP` (positive).
  `IRD`: Input file unit (opened in the main program).
  `IWR`: Output file unit (opened in the main program).

  o Output arguments:
  `NMAT`: Number of different materials in full bodies (excluding void regions).
  `NBOD`: Number of defined bodies and modules.

  Subroutine `GEOMIN` labels elements of the various kinds (surfaces, bodies and modules) in strictly increasing order; it may also redefine some of the geometry parameters, whose actual values are entered through the array `PARINP`. A copy of the geometry definition file, with the effective parameter values and with the element labels assigned by `GEOMIN`, is printed on the output file (`UNIT IWR`). This part of the output file describes the actual geometry used in the simulation.

- `SUBROUTINE LOCATE`
  Determines the body that contains the point with coordinates (`X`, `Y`, `Z`).

  o Input values (through `COMMON/TRACK/`) [4] :
  `X, Y, Z`: Particle position coordinates.
  `U, V, W`: Direction cosines of the direction of movement.
  o Output values (through `COMMON/TRACK/`):
  `IBODY`: Body where the particle moves.
  `MAT`: Material in `IBODY`. The output `MAT` = 0 indicates that the particle is in a void region.

- `SUBROUTINE STEP(DS,DSEF,NCROSS)`
  Used in conjunction with PENELOPE, this subroutine performs the geometrical part of the track simulation. The particle starts from the point (`X,Y,Z`) and proceeds to travel a length `DS` in the direction (`U,V,W`) within the material where it moves.

[4] Most of the input/output of the geometry routines is through `COMMON/TRACK/`, which is the common block used by PENELOPE to transfer particle state variables.

STEP displaces the particle and stops it at the end of the step, or just after entering a new material. The output value DSEF is the distance travelled within the initial material. If the particle enters a void region, STEP continues the particle track, as a straight segment, until it penetrates a material body or leaves the system (the path length through inner void regions is not included in DSEF). When the particle arrives from a void region (MAT = 0), it is stopped after entering the first material body. The output value MAT = 0 indicates that the particle has escaped from the system.

o Input-output values (through COMMON/TRACK/):
X, Y, Z: Input: coordinates of the initial position.
Output: coordinates of the final position.
U, V, W: Direction cosines of the displacement. They are kept unaltered.
IBODY Input: initial body, i.e. the one that contains the initial position.
Output: final body.
MAT: Material in body IBODY (automatically changed when the particle crosses an interface).

o Input argument:
DS: Distance to travel (unaltered).

o Output arguments:
DSEF: Travelled path length before leaving the initial material or completing the jump (less than DS if the track crosses an interface).
NCROSS: Number of interface crossings (=0 if the particle does not leave the initial material, greater than 0 if the particle enters a new material).

For the handling and storage of geometric information we take advantage of the structure of the genealogical tree. It is assumed that an enclosure has been defined so that it is the only common ancestor for all bodies and modules. To understand the operation of the geometry routines, it is convenient to define a matrix FLAG(KB,KS) as follows (the indices KS and KB indicate the label of a surface and a body or module, respectively),

FLAG(KB,KS) = 1, if KS is a limiting surface of KB and KB is inside KS (i.e. side pointer = -1).
= 2, if KS is a limiting surface of KB and KB is outside KS (i.e. side pointer = $+1$).
= 3, if KB is a body and KS does not directly limit KB, but appears in the definition of a body that limits KB.
= 4, if KB is a module and KS limits one of its daughters (bodies and submodules), but does not appear in the definition of KB.
= 5, otherwise.

To locate a point we call subroutine LOCATE, where we proceed upwards in the genealogical tree of modules. If the point is outside the enclosure, we set MAT = 0 and return to the main program. Otherwise, we look for a module or body of the second generation that contains the point. If it exists, we continue analyzing its descendants

(if any) and so on. The process ends when we have determined the body IBODY that contains the point, or as soon as we conclude that the point is outside the material system (i.e. in a void region). Notice that, when we have found that a module KB does contain the point, to do the next step we only need to consider the surfaces KS such that FLAG(KB,KS) = 1, 2 or 4.

After the body IBODY that contains the initial position of the particle has been identified, we can call subroutine STEP to move the particle a certain distance DS, dictated by PENELOPE, along the direction (U,V,W). We start by checking whether the track segment crosses any of the surfaces that limit IBODY. If after travelling the distance DS the particle remains within the same body, DSEF is set equal to DS and control is returned to the main program. It is worth noting that the surfaces KS that define the initial body are those with FLAG(IBODY,KS)=1 and 2 (proper limiting surfaces) or =3 (limiting surfaces of limiting bodies). Although it may happen that a surface with FLAG=3 does not directly limit the body, subroutine STEP cannot know this from the information at hand and, consequently, all surfaces with FLAG=3 are analyzed after each move. It is clear that, to reduce the number of surfaces to be considered, we should minimize the number of bodies used to delimit other bodies.

When the particle leaves IBODY and enters a new material, STEP stops it just after crossing the interface and determines the new body and material (in this case, the output values of IBODY and MAT are different from the input ones). To do this, the limiting surfaces of the parent module and of all the sisters of the initial body must be analyzed (if they exist). If the new position is outside the parent module, we must analyze all surfaces that limit the parent's sisters and go downward in the genealogical tree to determine the module that contains the point and, if necessary, go upwards again to find out what the new body is. If the new material is the same as in the initial body, the particle is allowed to move the remaining distance. Void regions (strict vacuum) are crossed freely (i.e. the distance travelled within these regions is not counted). Furthermore, when the particle starts from outside the enclosure, it is allowed to propagate freely until it reaches a material body. The particle is stopped when it penetrates a different material or when it leaves the system (i.e. when, after leaving a material body, its straight trajectory does not intersect a non-void body; in this case, the value MAT=0 is returned). Evidently, the speed of the geometry subroutines depends greatly on the structure of the modules' genealogical tree. The responsibility of optimizing it rests with the user.

When STEP moves the particle across an interface, there is a risk that, owing to numerical truncation errors, the particle is placed on the wrong side of the interface (i.e. the track is stopped just before the interface). If this occurs, the program could go into an endless loop in which STEP repeatedly tries to move the particle a very small distance (of the order of 10^{-15} cm) towards the interface but does not succeed, i.e. the particle is trapped at the interface. To avoid this collapse of the trajectory, after each interface crossing, STEP applies an additional small displacement ($\sim 10^{-8}$ cm) in the direction of movement, which is physically irrelevant and sufficient to compensate for the effect of truncation errors. The same strategy is used in subroutine LOCATE: when the particle is too close to an interface, it is moved 10^{-8} cm along the surface gradient direction

or its opposite, depending on whether the particle approaches or leaves the interface. Notice that this strategy requires that the direction of movement (U,V,W) be defined before calling LOCATE. The extra displacement effectively eliminates the risk of particle trapping at interfaces; but it also sets a limit to the space resolution (geometrical details that are less than ~10 Å in size cannot be described).

PENGEOM admits up to 250 surfaces and 125 bodies and modules. When the input file contains a larger number of elements, the program stops and a corresponding error message is printed. To describe such complex material systems, it is necessary to edit the source file PENGEOM.F and increase the values of the parameters NS (maximum number of surfaces) and NB (maximum number of bodies) in all subroutines. It is assumed that the number of bodies in a module is less than NX = 100, which is also the upper limit for the number of surfaces that can be used to define a body or a module (those with FLAG < 5). When NX is too small, the module that causes the trouble should be decomposed into several submodules. Although it is possible to increase the parameter NX, this would waste a lot of memory. As a consequence, a system with more than 100 surfaces or bodies must be decomposed into modules.

5.6 Debugging and viewing the geometry

A pair of computer programs named GVIEW2D and GVIEW3D have been written to visualize the geometry and to help the user to debug the definition file. These codes generate two- and three-dimensional 24-bit colour images of the system using specific graphics routines. The executable codes included in the distribution package run on personal computers under Microsoft Windows 9x.

The most characteristic (and useful) feature of GVIEW2D is that displayed pictures are generated by using the PENGEOM package and, therefore, errors and inconsistencies in the geometry definition file that would affect the results of actual simulations are readily identified. The method to generate the image consists of following a particle that moves on a plane perpendicular to an axis of the reference frame, which is mapped on the window. The particle starts from a position that corresponds to the left-most pixel and moves along a straight trajectory to the right of the window. To do this, we call subroutine STEP repeatedly, maintaining the direction of movement and with a large value of DS (such that each body is crossed in a single step). A colour code is assigned to each material, and pixels are lit up with the active colour when they are crossed by the particle trajectory. The active colour is changed when the particle enters a new material. The final picture is a map of the bodies and materials intersected by the window plane. The orientation of the window plane, as well as the position and size of the window view, may be changed interactively by entering one of the one-character commands shown in table 5.2, directly from the graphics window (upper- and lower-case letters may work differently). With GVIEW2D we can inspect the internal structure of the system with arbitrary magnification (limited only by the intrinsic resolution of the PENGEOM routines).

Table 5.2: One-character commands of the GVIEW2D geometry viewer.

```
++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++
+  x --> change window orientation, x-axis,                    +
+  y --> change window orientation, y-axis,                    +
+  z --> change window orientation, z-axis,                    +
+  r,right --> shift right,       l,left  --> shift left,      +
+  u,up    --> shift up,          d,down  --> shift down,      +
+  f,pgup  --> shift front,       b,pgdn  --> shift back,      +
+  i,+     --> zoom in,           o,-     --> zoom out,        +
+  1       --> actual size,       h,?     --> help,            +
+  blank, enter --> repeat last command,      q --> quit.      +
++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++
```

When running the GVIEW2D program, you will be asked to give the path+name of the geometry definition file and the coordinates (XC,YC,ZC) of the centre of the window (relative to the laboratory frame) in cm. The window may appear black (the colour for void regions) if no material bodies are intersected. In this case, use the one-character viewer commands to reach the bodies or, more conveniently, start again and place the window centre near or within a filled body.

GVIEW3D generates three-dimensional pictures of the geometry by using a simple ray-tracing algorithm, with the source light and the camera at the same position. Bodies are displayed with the same colour code used by GVIEW2D and the intensity of each pixel is determined by the angle between the vision line and the normal to the limiting surface. This method does not produce shadows and disregards light diffusion, but makes fairly realistic three-dimensional images. The camera is assumed to be outside the system (placing the camera inside a body would stop the program). To reveal the inner structure of the system, the program can eliminate a wedge (limited by two vertical planes that intersect in the z-axis). The position and size of the system can be modified by means of one-character commands entered from the graphics window. The command keys and actions are similar to those of GVIEW2D. It is worth noting that GVIEW3D generates the image pixel by pixel, whereas GVIEW2D does it by drawing straight lines on the window; as a result, GVIEW2D is much faster.

GVIEW2D and GVIEW3D produce an output file named GEOMETRY.REP (which is generated by subroutine GEOMIN) in the working directory. The programs are stopped either when an input format is incorrect (reading error) or when a clear inconsistency in the definition file is found (e.g. when the element that is being defined and the furnished information do not match). The wrong datum appears in the last printed lines of the GEOMETRY.REP file, usually in the last one. Error messages are also written on that file, so that the identification of inconsistencies is normally very easy. When the structure of the input file is correct, the codes do not stop and the geometry is displayed for further analysis. Most of the possible errors in the input file can only be revealed by direct inspection of the images generated by GVIEW2D and GVIEW3D.

The file GEOMETRY.REP is a duplicate of the input definition file. The only differ-

ences between the two files are the labels assigned to the different surfaces, bodies and modules; in GEOMETRY.REP, these elements are numbered in strictly increasing order. It is important to bear in mind that PENGEOM internally uses this sequential labelling to identify bodies and surfaces. Knowing the internal label assigned to each element is necessary for scoring purposes, e.g. to determine the distribution of energy deposited within a particular body.

5.7 A short tutorial

To prepare a new geometry definition file, it is useful to start from a file that contains a model of each data set with default values of their parameters. Placing the end-line at the beginning of the model group discontinues the geometry reading; so that the model group can be kept in the geometry file, even when this one is operative. The starting file should look like this

```
END      0000000000000000000000000000000000000000000000000000000000
0000000000000000000000000000000000000000000000000000000000000000
SURFACE (   )   REDUCED FORM
INDICES=( 1, 1, 1, 1, 1)
X-SCALE=(+1.000000000000000E+00,  0)              (DEFAULT=1.0)
Y-SCALE=(+1.000000000000000E+00,  0)              (DEFAULT=1.0)
Z-SCALE=(+1.000000000000000E+00,  0)              (DEFAULT=1.0)
  OMEGA=(+0.000000000000000E+00,  0) DEG          (DEFAULT=0.0)
  THETA=(+0.000000000000000E+00,  0) DEG          (DEFAULT=0.0)
    PHI=(+0.000000000000000E+00,  0) RAD          (DEFAULT=0.0)
X-SHIFT=(+0.000000000000000E+00,  0)              (DEFAULT=0.0)
Y-SHIFT=(+0.000000000000000E+00,  0)              (DEFAULT=0.0)
Z-SHIFT=(+0.000000000000000E+00,  0)              (DEFAULT=0.0)
0000000000000000000000000000000000000000000000000000000000000000
SURFACE (   )   IMPLICIT FORM
INDICES=( 0, 0, 0, 0, 0)
    AXX=(+0.000000000000000E+00,  0)              (DEFAULT=0.0)
    AXY=(+0.000000000000000E+00,  0)              (DEFAULT=0.0)
    AXZ=(+0.000000000000000E+00,  0)              (DEFAULT=0.0)
    AYY=(+0.000000000000000E+00,  0)              (DEFAULT=0.0)
    AYZ=(+0.000000000000000E+00,  0)              (DEFAULT=0.0)
    AZZ=(+0.000000000000000E+00,  0)              (DEFAULT=0.0)
     AX=(+0.000000000000000E+00,  0)              (DEFAULT=0.0)
     AY=(+0.000000000000000E+00,  0)              (DEFAULT=0.0)
     AZ=(+0.000000000000000E+00,  0)              (DEFAULT=0.0)
     A0=(+0.000000000000000E+00,  0)              (DEFAULT=0.0)
0000000000000000000000000000000000000000000000000000000000000000
BODY    (   )   TEXT
MATERIAL(   )
SURFACE (   ), SIDE POINTER=( 1)
BODY    (   )
0000000000000000000000000000000000000000000000000000000000000000
MODULE  (   )   TEXT
MATERIAL(   )
SURFACE (   ), SIDE POINTER=( 1)
BODY    (   )
```


 363

```
MODULE  (   )
1111111111111111111111111111111111111111111111111111111111111111
  OMEGA=(+0.000000000000000E+00,  0) DEG          (DEFAULT=0.0)
  THETA=(+0.000000000000000E+00,  0) DEG          (DEFAULT=0.0)
    PHI=(+0.000000000000000E+00,  0) RAD          (DEFAULT=0.0)
X-SHIFT=(+0.000000000000000E+00,  0)              (DEFAULT=0.0)
Y-SHIFT=(+0.000000000000000E+00,  0)              (DEFAULT=0.0)
Z-SHIFT=(+0.000000000000000E+00,  0)              (DEFAULT=0.0)
0000000000000000000000000000000000000000000000000000000000000000
```

Then, to generate a new element, we just duplicate the corresponding data set, modify the parameter values and eliminate the lines that are unnecessary (i.e. those of parameters that take their default values). Of course, the defining data set must be placed before the end-line. The progressing geometry can be visualized with GVIEW2D as soon as the first complete body has been defined. If GVIEW2D stops before entering the graphics mode, the geometry definition is incorrect and we should have a look at the `GEOMETRY.REP` file to identify the problem. Normally, the conflicting parameter or element appears in the last line of this file.

The basic elements of the geometry definition are quadric surfaces. These can be visualized by using the following simple file, which defines the inner volume of a reduced quadric as a single body,

```
------------------------------------------------------------
        Visualization of reduced quadric surfaces.
Define the desired quadric (surface 1) by entering its indices.
The region with side pointer -1 (inside the quadric) corresponds
to MATERIAL=1.
0000000000000000000000000000000000000000000000000000000000000000
SURFACE (  1)   Reduced quadric. One sheet hyperboloid.
INDICES=( 1, 1,-1, 0,-1)
0000000000000000000000000000000000000000000000000000000000000000
BODY    (  1)   The interior of the quadric.
MATERIAL(  1)
SURFACE (  1), SIDE POINTER=(-1)
0000000000000000000000000000000000000000000000000000000000000000
END     0000000000000000000000000000000000000000000000000000000000
```

Notice that, in this case, the body is infinite in extent. There is no objection to using infinite bodies, as long as the enclosure contains all material bodies. When only a single body is defined, PENGEOM identifies it as the enclosure, and this requirement is met. Otherwise, we must define a proper enclosure (since all bodies and modules must have a common ancestor).

The following example describes a sphere with an inner arrow (fig. 5.2):

```
XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX
  Sphere of 5 cm radius with an arrow.
0000000000000000000000000000000000000000000000000000000000000000
SURFACE (  1)   PLANE Z=4.25
INDICES=( 0, 0, 0, 1,-1)
Z-SCALE=( 4.250000000000000E+00,  0)
0000000000000000000000000000000000000000000000000000000000000000
SURFACE (  2)   PLANE Z=1.5
```

```
INDICES=( 0, 0, 0, 1,-1)
Z-SCALE=( 1.500000000000000E+00,  0)
0000000000000000000000000000000000000000000000000000000000000000000000
SURFACE (  3)   PLANE Z=-4.0
INDICES=( 0, 0, 0, 1, 1)
Z-SCALE=( 4.000000000000000E+00,  0)
0000000000000000000000000000000000000000000000000000000000000000000000
SURFACE (  4)   CONE
INDICES=( 1, 1,-1, 0, 0)
X-SCALE=( 5.000000000000000E-01,  0)
Y-SCALE=( 5.000000000000000E-01,  0)
Z-SHIFT=( 4.250000000000000E+00,  0)
0000000000000000000000000000000000000000000000000000000000000000000000
SURFACE (  5)   CYLINDER
INDICES=( 1, 1, 0, 0,-1)
X-SCALE=( 7.250000000000000E-01,  0)
Y-SCALE=( 7.250000000000000E-01,  0)
0000000000000000000000000000000000000000000000000000000000000000000000
BODY    (  1)   ARROW HEAD
MATERIAL(  2)
SURFACE (  1), SIDE POINTER=(-1)
SURFACE (  2), SIDE POINTER=( 1)
SURFACE (  4), SIDE POINTER=(-1)
0000000000000000000000000000000000000000000000000000000000000000000000
BODY    (  2)   ARROW STICK
MATERIAL(  2)
SURFACE (  5), SIDE POINTER=(-1)
SURFACE (  2), SIDE POINTER=(-1)
SURFACE (  3), SIDE POINTER=( 1)
0000000000000000000000000000000000000000000000000000000000000000000000
SURFACE (  6)   SPHERE. R=5
INDICES=( 1, 1, 1, 0,-1)
X-SCALE=( 5.000000000000000E+00,  0)
Y-SCALE=( 5.000000000000000E+00,  0)
Z-SCALE=( 5.000000000000000E+00,  0)
0000000000000000000000000000000000000000000000000000000000000000000000
MODULE  (  3)   SPHERE WITH INNER ARROW
MATERIAL(  1)
SURFACE (  6), SIDE POINTER=(-1)
BODY    (  1)
BODY    (  2)
1111111111111111111111111111111111111111111111111111111111111111111111
  OMEGA=( 0.000000000000000E+00,  0) DEG
  THETA=(-90.00000000000000E+00,  0) DEG
    PHI=( 90.00000000000000E+00,  0) DEG
0000000000000000000000000000000000000000000000000000000000000000000000
END     0000000000000000000000000000000000000000000000000000000000000000000000
```

We have defined the entire system as a single module, so that you may rotate and/or displace it arbitrarily. Notice that the initial arrow points in the positive direction of the z-axis. It is instructive to try various rotations and use GVIEW2D or GVIEW3D (with a sector excluded to make the inner arrow visible) for visualizing the rotated system.

Writing a geometry file is nothing more than routine work. After a little practice, you can define quite complex systems by using only surfaces and bodies. You will soon

realize that the visualization programs (as well as the actual simulations!) slow down when the number of elements in the geometry increases. The only way of speeding up the programs is to group the bodies into modules. The best strategy for improving the calculation speed is to build relatively simple modules and combine them into larger parent modules to obtain a genealogical tree where the number of daughters of each module is not too large (say 4 or 5).

You may save a lot of time by defining each body separately (and checking it carefully) and then inserting it into the progressing module that, once finished, will be added to the file. Notice that the input element labels are arbitrary (as long as they are not repeated for elements of the same kind) and that we can insert new elements anywhere in the file. Once the geometry definition is complete, we can generate an equivalent file, with elements labelled according to their input order, by simply editing the `GEOMETRY.REP` file.

The previous examples of geometry files (`QUADRIC` and `ARROW`) together with several other files of more complex geometries are included in the distribution package. They can be directly visualized by running GVIEW2D and GVIEW3D. The file `GLASS` (a glass of champagne) shows that common objects can be described quite precisely with only quadric surfaces; in this case, we do not use modules, which are useful only to accelerate the calculations. `WELL` defines a scintillation well detector with much detail; we have set an enclosure for the system, so that you can rotate the entire detector by editing the definition file. Notice that, when the detector is tilted, it is very difficult to get an idea of its geometry from the images generated by GVIEW2D. `SATURN` describes the head of an electron accelerator, quite a complicated geometry with 96 surfaces and 44 bodies. The structure `MALE`, which corresponds to a mathematical anthropomorphic phantom, consists of 174 surfaces and 108 bodies, grouped into 11 modules.

We cannot finish without a word of caution about the use of PENGEOM, and other general-purpose geometry packages. For simple geometries, they tend to waste a lot of time. It is always advisable to consider the possibility of handling geometric aspects directly; this may enable substantial reduction of the number of operations by taking full advantage of the peculiarities of the material system.

Chapter 6

Structure and operation of the code system

In this chapter we describe the structure of the PENELOPE code system and its operation. The kernel of the system is the FORTRAN77 subroutine package PENELOPE, which performs "analogue" simulation of electron-photon showers (i.e. the simulated showers are intended to be replicas of actual showers) in infinite (unbounded) media of various compositions. Photon histories are generated by using the detailed simulation method (see section 1.4), i.e. all interaction events are simulated in chronological succession. The generation of electron and positron tracks is performed by using the mixed procedure described in chapter 4. Secondary particles emitted with initial energy larger than the absorption energy –see below– are stored, and simulated after completion of each primary track. Secondary particles are produced in direct interactions (hard inelastic collisions, hard bremsstrahlung emission, positron annihilation, Compton scattering, photoelectric absorption and pair production) and as fluorescent radiation (characteristic x-rays and Auger electrons). PENELOPE simulates fluorescent radiation that results from vacancies produced in K-shells and L-subshells by photoelectric absorption and Compton scattering of photons and by electron/positron impact. The relaxation of these vacancies is followed until the K- and L-shells are filled up, i.e. until the vacancies have migrated to M and outer shells.

Being a subroutine package, PENELOPE cannot operate by itself. The user must provide a steering `MAIN` program for his/her particular problem. Nevertheless, this `MAIN` program is normally fairly simple, since it only has to control the evolution of the tracks simulated by PENELOPE and keep score of relevant quantities. PENELOPE is devised to do the largest part of the simulation work. It allows the user to write his or her own simulation program, with arbitrary geometry and scoring, without previous knowledge of the intricate theoretical aspects of scattering and transport theories. In the case of material systems with quadric geometries, the geometrical operations can be done automatically by using the package PENGEOM (see chapter 5). The distribution package also includes various examples of `MAIN` programs for simple geometries (slab and cylindrical) and for general quadric geometries with limited scoring. Although

they are mostly intended to illustrate the use of the simulation routines, they do allow studying many cases of practical interest. The complete program system is written in FORTRAN77 (ANSI/ISO standard form) and, therefore, it should run on any platform with a FORTRAN77 or FORTRAN90 compiler.

6.1 PENELOPE

PENELOPE simulates coupled electron-photon transport in arbitrary material systems consisting of a number of homogeneous regions (bodies) limited by sharp (and passive) interfaces. Initially, it was devised to simulate the PENetration and Energy LOss of Positrons and Electrons in matter; photons were introduced later. The adopted interaction models (chapters 2-4), and the associated databases, allow the simulation of electron/positron and photon transport in the energy range from 100 eV to 1 GeV.

It should be borne in mind that our approximate interaction models become less accurate when the energy of the transported radiation decreases. Actually, for energies below ~1 keV, the DCSs are not well known, mostly because they are strongly affected by the state of aggregation. On the other hand, for electrons and positrons, the trajectory picture ceases to be applicable (because coherent scattering from multiple centers becomes appreciable) when the de Broglie wavelength, $\lambda_B = (150 \text{ eV}/E)^{1/2}$ Å, is similar to or greater than the interatomic spacing (~ 1 Å). Therefore, results from simulations with PENELOPE (or with any other Monte Carlo trajectory code) for energies below 1 keV or so, should be considered to have only a qualitative (or, at most, semi-quantitative) value. We recall also that, for elements with intermediate and high atomic numbers, secondary characteristic photons with energies less than the M-shell absorption edge are not simulated by PENELOPE. This sets a lower limit to the energy range for which the simulation is faithful.

The source file `PENELOPE.F` (about 8000 lines of FORTRAN code) consists of four blocks of subprograms, namely, preparatory calculations and I/O routines, interaction simulation procedures, numerical routines and transport routines. Only the latter are invoked from the `MAIN` program. The interaction simulation routines implement the theory and algorithms described in chapters 2 and 3. Although the interaction routines are not called from the `MAIN` program, there are good reasons to have them properly identified. Firstly, these are the code pieces to be modified to incorporate better physics (when available) and, secondly, some of these subroutines deliver numerical values of the DCSs (which can be useful to apply certain variance reduction techniques). To have these routines organized, we have named them according to the following convention:

- The first letter indicates the particle (`E` for electrons, `P` for positrons, `G` for photons).
- The second and third letters denote the interaction mechanism (`EL` for elastic, `IN` for inelastic, `BR` for bremsstrahlung, `AN` for annihilation, `RA` for Rayleigh, `CO` for Compton, `PH` for photoelectric and `PP` for pair production).
- The random sampling routines have three-letter names. Auxiliary routines, which perform specific calculations, have longer names, with the fourth and subsequent letters

and/or numbers indicating the kind of calculation (TX for total x-section, DX for differential x-section) or action (W for write data on a file, R for read data from a file, I for initialization of simulation algorithm).
Thus, for instance, subroutine EEL simulates elastic collisions of electrons while subroutine EINTX computes total (integrated) cross sections for inelastic scattering of electrons.

6.1.1 Database and input material data file

PENELOPE reads the required physical information about each material (which includes tables of physical properties, interaction cross sections, relaxation data, etc.) from the input material data file (identified as UNIT=IRD in the code source listing). The material data file is created by means of the auxiliary program MATERIAL, which extracts atomic interaction data from the database. This program runs interactively and is self-explanatory. Basic information about the considered material is supplied by the user from the keyboard, in response to prompts from the program. The required information is: 1) chemical composition (i.e. elements present and stoichiometric index of each element), 2) mass density, 3) mean excitation energy and 4) energy and oscillator strength of plasmon excitations. Alternatively, for a set of 279 prepared materials, the program MATERIAL can read data directly from the PDCOMPOS.TAB file (see below).

For compounds and mixtures, the additivity approximation is adopted to define the material's cross sections, i.e. the corresponding "molecular" cross section is set equal to the sum of atomic cross sections weighted with the stoichiometric index of the element. Alloys and mixtures are treated as compounds, with stoichiometric indices equal, or proportional, to the percent number of atoms of the elements.

The PENELOPE database consists of the following 373 ASCII files,

PDATCONF.TAB ... Atomic ground-state configurations, ionization energies (Lederer and Shirley, 1978) and central values, $J_i(p_z = 0)$, of the one-electron shell Compton profiles (Biggs et al., 1975) for the elements, from hydrogen to uranium.

PDCOMPOS.TAB ... This file contains composition data, densities and mean excitation energies for 279 materials, adapted from the database of the ESTAR program of Berger (1992). The first 98 entries are the elements $Z = 1 - 98$, ordered by atomic number Z. Materials 99 to 279 are compounds and mixtures, in alphabetical order. Notice that PENELOPE does not work for elements with atomic number $Z > 92$.

PDEFLIST.TAB ... List of materials predefined in file PDCOMPOS.TAB, with their identification numbers.

PDRELAX.TAB ... Data on atomic relaxation, extracted from the LLNL Evaluated Atomic Data Library (Perkins et al., 1991)

92 files named PDEELZZ.TAB with ZZ=atomic number (01–92). These files contain integrated cross sections for elastic scattering of electrons and positrons by neutral

 3610

atoms, calculated by using the partial-wave methods described in section 3.1 (Salvat, 2000). The first line in each file gives the atomic number ZZ; each subsequent line has 7 columns with the following data:
1st column: kinetic energy (eV), in increasing order.
2nd column: total cross section for electrons.
3rd column: first transport cross section for electrons.
4th column: second transport cross section for electrons.
5th column: total cross section for positrons.
6th column: first transport cross section for positrons.
7th column: second transport cross section for positrons.
The grid of energies is approximately logarithmic, with 15 points per decade, and is the same for all elements. All cross sections are in cm^2.

92 files named PDEBRZZ.TAB with ZZ=atomic number (01–92). They contain the atomic bremsstrahlung scaled cross sections (energy loss spectra) and total integrated radiative cross sections of electrons, for a grid of electron kinetic energies E and reduced photon energies W/E that is dense enough to allow the use of cubic spline log-log interpolation in E and linear interpolation in W/E. The data in these files is from a database, with 32 reduced photon energies, which was provided to the authors by Steven Seltzer (a brief description of the methods used to compute the database and a reduced tabulation is given in Seltzer and Berger, 1986). The format of the bremsstrahlung database files is the following,
1) The first line contains the atomic number ZZ.
2) Each four-lines block contains the electron kinetic energy E, the scaled energy-loss differential cross section at the 32 fixed reduced photon energies and the value of the integrated radiative cross section.
Energies are in eV and the values of the scaled energy-loss cross section are in millibarn (10^{-27} cm^2).

PDBRANG.TAB ... Gives the parameters of the analytical shape function (angular distribution) of bremsstrahlung photons, which is expressed as a statistical mixture of two Lorentz-boosted dipole distributions, eq. (3.146). The distribution parameters were obtained by fitting the benchmark partial-wave shapes tabulated by Kissel et al. (1983).

92 files named PDGPPZZ.TAB with ZZ=atomic number (01–92). Total cross sections for electron-positron pair production by photons with energies up to 100 GeV in the field of neutral atoms. The data were generated by means of the XCOM program of Berger and Hubbell (1987). The first line of each file gives the atomic number ZZ; each subsequent line gives,
1st column: photon energy, in eV. The same energy grid for all elements.
2nd column: total cross section for pair+triplet production in barn (10^{-24} cm^2).

92 files named PDGPHZZ.TAB with ZZ=atomic number (01–92), containing photoelectric total atomic cross sections and partial cross sections for photoionization of inner shells (K shell and L subshells) for the elements and photon energies in the

range from 100 eV to 1 TeV. The data were extracted from the LLNL Evaluated Photon Data Library EPDL97 (Cullen et al., 1997). The format is the following,
1) the first line contains the atomic number ZZ and the number NS of shells for which the partial cross section is tabulated.
2) each of the following lines contains a value of the photon energy (in eV) and the corresponding total cross section and partial cross sections of the shells K, L1, L2 and L3, respectively (all cross sections in barn). For low-Z elements, L-subshells are empty and, therefore, they do not appear in the table.
The grid of energies for each element was obtained by merging a generic grid (the same for all elements, covering the energy range from 100 eV to 100 GeV) with the grid of absorption edges of the element, and adding additional points (where needed) to ensure that linear log-log interpolation will never introduce relative errors larger than 0.02.

Atomic cross sections for coherent and incoherent scattering of photons, inelastic scattering of electrons and positrons, and positron annihilation are evaluated directly from the analytical DCSs described in chapters 2 and 3 .

In the material definition file generated by the program MATERIAL, mean free paths, transport mean free paths and stopping powers of electrons and positrons are given in mass-thickness units (1 mtu $\equiv$ 1 g/cm^2) and eV/mtu, respectively. Photon mass attenuation coefficients are expressed in cm^2/g. These quantities are practically independent of the material density; the only exception is the collision stopping power for electrons and positrons with kinetic energies larger than about 0.5 MeV, for which the density effect correction may be appreciable.

The energy-dependent quantities tabulated in the input material data file determine the most relevant characteristics of the scattering model. Thus, the MW differential cross section for electron and positron elastic scattering is completely defined by the mean free paths and transport mean free paths. Collision and radiative stopping powers read from the input file are used to renormalize the built-in analytical differential cross sections, i.e. these are multiplied by an energy-dependent factor such that the input stopping powers are exactly reproduced. The mean free paths used in the simulation of photon transport are directly obtained from the input total cross sections. Natural cubic spline log-log interpolation is used to interpolate the tabulated energy-dependent quantities, except for the photoelectric attenuation coefficient, which is obtained by simple linear log-log interpolation in the intervals between consecutive absorption edges.

To simulate geometrical structures with several materials, the corresponding material data files generated by the program MATERIAL must be catenated in a single input file. PENELOPE labels the M-th material in this file with the index MAT=M, which is used during the simulation to identify the material where the particle moves. The maximum number of different materials that PENELOPE can handle simultaneously is fixed by the parameter MAXMAT, which in the present version is set equal to 10. The required memory storage is roughly proportional to the value of this parameter. The user can increase MAXMAT by editing the program source files. Notice that the value of MAXMAT *must* be

the same in all subprograms.

6.1.2 Structure of the MAIN program

As mentioned above, PENELOPE must be complemented with a steering MAIN program, which controls the geometry and the evolution of tracks, keeps score of the relevant quantities and performs the required averages at the end of the simulation.

The connection of PENELOPE and the MAIN program is done via the named common block

→ `COMMON/TRACK/E,X,Y,Z,U,V,W,WGHT,KPAR,IBODY,MAT,ILB(5)`

that contains the following particle state variables:

`KPAR` ... kind of particle (1: electron, 2: photon, 3: positron).

`E` ... current particle energy (eV) (kinetic energy for electrons and positrons).

`X, Y, Z` ... position coordinates (cm).

`U, V, W` ... direction cosines of the direction of movement.

`WGHT` ... in analogue simulations, this is a dummy variable. When using variance reduction methods, the particle weight can be stored here.

`IBODY` ... this auxiliary flag serves to identify different bodies in complex material structures.

`MAT` ... material where the particle moves (i.e. the one in the body labelled `IBODY`).

`ILB(5)` ... an auxiliary array of 5 labels that describe the origin of secondary particles (see below). It is useful e.g. to study partial contributions from particles originated by a given process.

The position coordinates $\mathbf{r}$ =(X,Y,Z) and the direction cosines $\hat{\mathbf{d}}$ =(U,V,W) of the direction of movement are referred to a fixed rectangular coordinate system, the "laboratory" system, which can be arbitrarily defined. During the simulation, all energies and lengths are expressed in eV and cm, respectively.

The label `KPAR` identifies the kind of particle: `KPAR=1`, electron; `KPAR=2`, photon; `KPAR=3`, positron. A particle that moves in material `M` is assumed to be absorbed when its energy becomes less than a value `EABS(KPAR,M)` (in eV) specified by the user. Positrons are assumed to annihilate, by emission of two photons, when absorbed. In dose calculations, `EABS(KPAR,M)` should be determined so that the residual range of particles with this energy is smaller than the dimensions of the volume bins used to tally the spatial dose distribution. As the interaction database is limited to energies above 100 eV, absorption energies `EABS(KPAR,M)` must be larger than this value.

The transport algorithm for electrons and positrons in each material `M` is controlled by the following simulation parameters,

C1(M) ... Average angular deflection, $C_1 \simeq 1 - \langle \cos\theta \rangle$ [eq. (4.11)], produced by multiple elastic scattering along a path length equal to the mean free path between consecutive hard elastic events [see eq. (4.1)]. C1(M) should be of the order of 0.05; its maximum allowed value is 0.2.

C2(M) ... Maximum average fractional energy loss, C_2 [eq. (4.85)], between consecutive hard elastic events. Usually, a value of the order of 0.05 is adequate. The maximum allowed value of C2(M) is 0.2.

WCC(M) ... Cutoff energy loss, W_{cc} (in eV), for hard inelastic collisions in the Mth material.

WCR(M) ... Cutoff energy loss, W_{cr} (in eV), for hard bremsstrahlung emission in material M.

These parameters determine the accuracy and speed of the simulation. To ensure accuracy, C1(M) and C2(M) should have small values (of the order of 0.01 or so). With larger values of C1(M) and C2(M) the simulation gets faster, at the expense of a certain loss in accuracy. The cutoff energies WCC(M) and WCR(M) mainly influence the simulated energy distributions. The simulation speeds up by using larger cutoff energies, but if these are too large, the simulated energy distributions may be somewhat distorted. In practice, simulated energy distributions are found to be insensitive to the adopted values of WCC(M) and WCR(M) when these are less than the bin width used to tally the energy distributions. Thus, the desired energy resolution determines the maximum allowed cutoff energies. The reliability of the whole simulation rests on a single condition: the number of steps (or random hinges) per primary track must be "statistically sufficient", i.e. larger than 10 or so.

The simulation package is initialized from the MAIN program with the statement

```
→ CALL PEINIT(EPMAX,NMAT,IRD,IWR,INFO)
```

Subroutine PEINIT reads the data files of the different materials, evaluates relevant scattering properties and prepares look-up tables of energy-dependent quantities that are used during the simulation. Its input arguments are:

EPMAX ... Maximum energy (in eV) of the simulated particles. Notice that if the primary particles are positrons with initial kinetic energy EP, the maximum energy of annihilation photons may be close to (but less than) EPMAX $= 1.21$(EP $+ m_e c^2$); in this special case, the maximum energy is larger than the initial kinetic energy.

NMAT ... Number of different materials (less than or equal to MAXMAT).

IRD ... Input unit.

IWR ... Output unit.

INFO ... Determines the amount of information that is written on the output unit. Minimal for INFO=0 and increasingly detailed for INFO=1, 2, etc.

For the preliminary computations, PEINIT needs to know the absorption energies EABS(KPAR,M) and the simulation parameters C1(M), C2(M), WCC(M) and WCR(M). This information is introduced through the named common block

```
→  COMMON/CSIMPA/EABS(3,MAXMAT),C1(MAXMAT),C2(MAXMAT),WCC(MAXMAT),
   1  WCR(MAXMAT)
```

that has to be loaded before invoking subroutine PEINIT. Notice that we can employ different values of the simulation parameters for different materials. This possibility can be used to speed up the simulation in regions of lesser interest.

PENELOPE has been structured in such a way that a particle track is generated as a sequence of track segments (free flights or "jumps"); at the end of each segment, the particle suffers an interaction with the medium (a "knock") where it loses energy, changes its direction of movement and, in certain cases, produces secondary particles. Electron-photon showers are simulated by successively calling the following subroutines:

SUBROUTINE CLEANS ... Initiates the secondary stack.

SUBROUTINE START ... For electrons and positrons, this subroutine forces the following interaction event to be a soft artificial one. It must be called before starting a new –primary or secondary– track and also when a track crosses an interface.
Calling START is strictly necessary only for electrons and positrons; for photons this subroutine has no physical effect. However, it is advisable to call START for any kind of particle since it checks whether the energy is within the expected range, and can thus help to detect "bugs" in the MAIN program.

SUBROUTINE JUMP(DSMAX,DS) ... Determines the length DS of the track segment to the following interaction event.
The input parameter DSMAX defines the maximum allowed step length for electrons/positrons; for photons, it has no effect. As mentioned above, to limit the step length, PENELOPE places delta interactions along the particle track. These are fictitious interactions that do not alter the physical state of the particle. Their only effect is to interrupt the sequence of simulation operations (which requires altering the values of inner control variables to permit resuming the simulation in a consistent way). The combined effect of the soft interactions that occur along the step preceding the delta interaction is simulated by the usual random hinge method.
As mentioned above, to ensure the reliability of the mixed simulation algorithm, the number of artificial soft events per particle track in each body should be larger than, say, 10. For relatively thick bodies (say, thicker than 10 times the mean free path between hard interactions), this condition is automatically satisfied. In this case we can switch off the step-length control by setting DSMAX=1.0D35 (or any other very large value). On the other hand, when the particle moves in a thin body, DSMAX should be given a value of the order of one tenth of the "thickness" of that body. Limiting the step length is also necessary to simulate particle transport in external electromagnetic fields.

SUBROUTINE KNOCK(DE,ICOL) ... Simulates an interaction event, computes new energy and direction of movement, and stores the initial states of the generated secondary particles, if any. On output, the arguments are:
DE ... deposited energy in the course of the event,
ICOL ... kind of event that has been simulated, according to the following convention,

- Electrons (KPAR=1)
  ICOL=1, artificial soft event (random hinge).
  =2, hard elastic collision.
  =3, hard inelastic collision.
  =4, hard bremsstrahlung emission.
- Photons (KPAR=2)
  ICOL=1, coherent (Rayleigh) scattering.
  =2, incoherent (Compton) scattering.
  =3, photoelectric absorption.
  =4, electron-positron pair production.
- Positrons (KPAR=3)
  ICOL=1, artificial soft event (random hinge).
  =2, hard elastic collision.
  =3, hard inelastic collision.
  =4, hard bremsstrahlung emission.
  =5, annihilation.

For electrons and positrons ICOL=7 corresponds to delta interactions. The value ICOL=6 is used for the "auxiliary" interactions (an additional mechanism that may be defined by the user, e.g. to simulate photonuclear interactions, see the source file PENELOPE.F).

SUBROUTINE SECPAR(LEFT) ... Sets the initial state of a secondary particle and removes it from the secondary stack. The output value LEFT is the number of secondary particles that remained in the stack at the calling time.

SUBROUTINE STORES(E,X,Y,Z,U,V,W,WGHT,KPAR,ILB) ... Stores a particle in the secondary stack. Arguments have the same meaning as in COMMON/TRACK/, but refer to the particle that is being stored. The variables IBODY and MAT are set equal to the current values in COMMON/TRACK/.
Calling STORES from the MAIN program is useful e.g. to store particles produced by splitting, a variance reduction method (see section 1.6.2).

The sequence of calls to generate a random track is independent of the kind of particle that is being simulated. The generation of random showers proceeds as follows (see fig. 6.1):

(i) Set the initial state of the primary particle, i.e. assign values to the state variables KPAR, E, position coordinates $\mathbf{r}$ =(X,Y,Z) and direction of movement $\hat{\mathbf{d}}$ =(U,V,W). Specify the body and material where the particle moves by defining the values of IBODY and MAT, respectively. Optionally, set the values of WGHT and ILB(1:5).

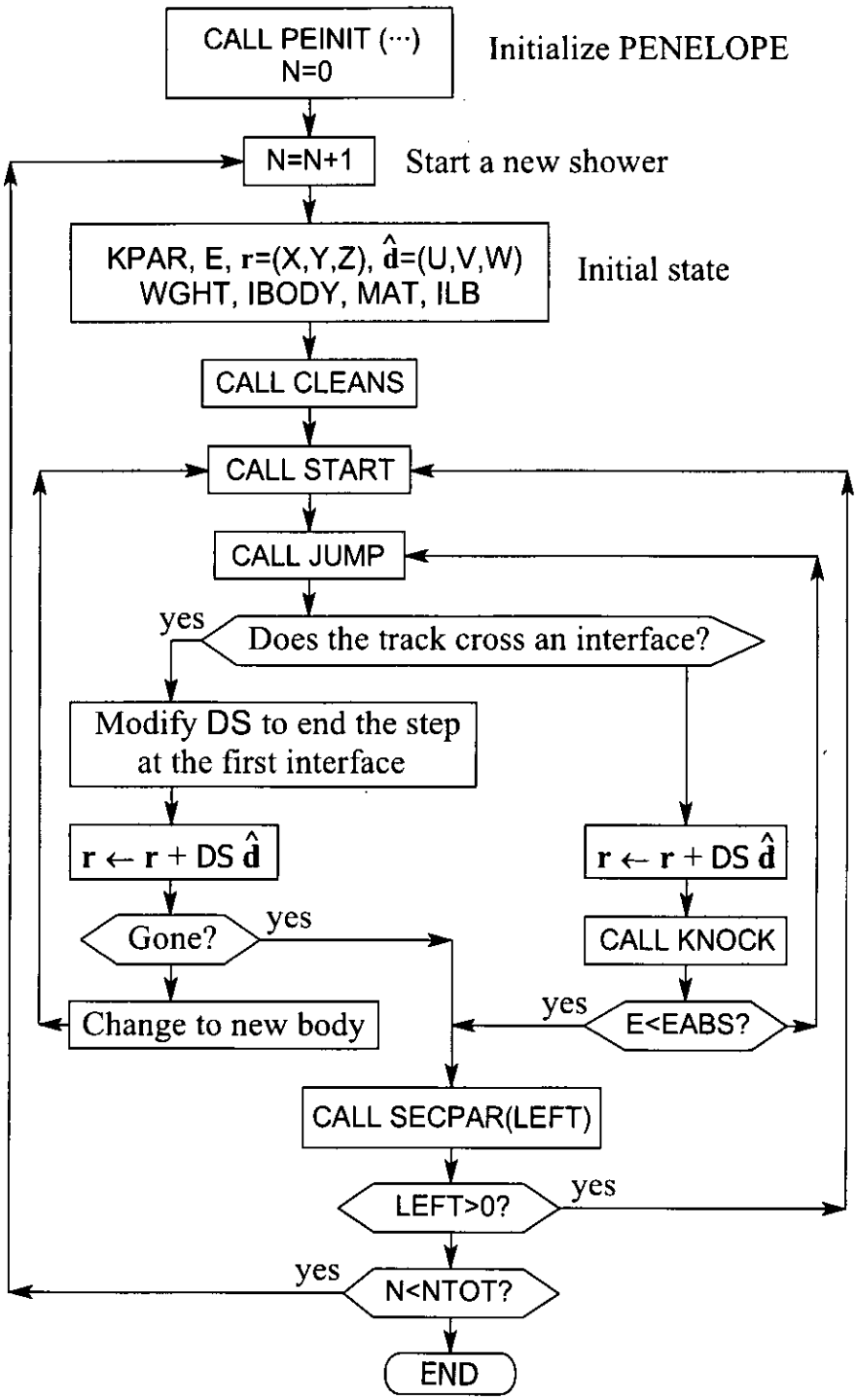


Figure 6.1: Flow diagram of the `MAIN` program for simulating electron-photon showers with PENELOPE.

(ii) CALL CLEANS to initialize the secondary stack.

(iii) CALL START to initiate the simulation of the track.

(iv) CALL JUMP(DSMAX,DS) to determine the length DS of the next track segment (for electrons and positrons, DS will never exceed the input value DSMAX).

(v) Compute the position of the following event:

- If the track has crossed an interface, stop the particle at the position where the track intersects the interface, and shorten the step length DS accordingly. Change to the new material (the one behind the interface) by redefining the variables IBODY and MAT.
  When the particle escapes from the system, the simulation of the track has been finished; increment counters and go to step (vii).
  Go to step (iii).

(vi) CALL KNOCK(DE,ICOL) to simulate the following event.

- If the energy is less than EABS(KPAR,MAT), end the track, increment counters and go to step (vii).
- Go to step (iv).

(vii) CALL SECPAR(LEFT) to start the track of a particle in the secondary stack (this particle is then automatically removed from the stack).

- If LEFT>0, go to step (iii). The initial state of a secondary particle has already been set.
- If LEFT=0, the simulation of the shower produced by the primary particle has been completed. Go to step (i) to generate a new primary particle (or leave the simulation loop after simulating a sufficiently large number of showers).

Notice that subroutines JUMP and KNOCK keep the position coordinates unaltered; the positions of successive events have to be followed by the MAIN program (simply by performing a displacement of length DS along the direction of movement after each call to JUMP). The energy of the particle is automatically reduced by subroutine KNOCK, after generating the energy loss from the relevant probability distribution. KNOCK also modifies the direction of movement according to the scattering angles of the simulated event. Thus, at the output of KNOCK, the values of the energy E, the position $\mathbf{r}$ =(X,Y,Z) and the direction of movement $\hat{\mathbf{d}}$ =(U,V,W) define the particle state immediately after the interaction event.

In order to avoid problems related with possible overflows of the secondary stack, when a secondary particle is produced its energy is temporarily assumed as locally deposited. Hence, the energy E of a secondary must be subtracted from the corresponding dose counter when the secondary track is started. Occasional overflows of the secondary stack are remedied by eliminating the less energetic secondary electron or photon in the stack (positrons are not eliminated since they will eventually produce quite energetic annihilation radiation). As the main effect of secondary particles is to spread out the

energy deposited by the primary one, the elimination of the less energetic secondary electrons and photons should not invalidate local dose calculations.

It is the responsibility of the user to avoid calling subroutines JUMP and KNOCK with energies outside the interval (EABS(KPAR,M),EMAX). This could cause improper interpolation of the cross sections. The simulation is aborted (and an error message is printed in unit 6) if the conditions EABS(KPAR)<E<EMAX are not satisfied when a primary or secondary track is started (whenever subroutine START is called at the beginning of the track).

Pseudo-random numbers uniformly distributed in the interval (0,1) are supplied by function RAND(DUMMY) that implements a 32-bit generator due to L'Ecuyer (see table 1.1). The seeds of the generator (two integers) are transferred from the MAIN program through the named common block RSEED (see below). The random number generator can be changed by merely replacing that FUNCTION subprogram (the new one has to have a single dummy argument). Some compilers incorporate an intrinsic random number generator with the same name (but with different argument lists). To avoid conflict, RAND should be declared as an external function in all subprograms that call it.

Notice that

(1) In the simulation routines, real and integer variables are declared as DOUBLE PRECISION and INTEGER*4, respectively. To prevent type mismatches, it is prudent to use the following IMPLICIT statement

```
→ IMPLICIT DOUBLE PRECISION (A-H,O-Z), INTEGER*4 (I-N)
```

in the MAIN program and other user program units.

(2) The MAIN program *must* include the following three common blocks:

```
→ COMMON/TRACK/E,X,Y,Z,U,V,W,WGHT,KPAR,IBODY,MAT,ILB(5)
→ COMMON/CSIMPA/EABS(3,MAXMAT),C1(MAXMAT),C2(MAXMAT),WCC(MAXMAT),
  1  WCR(MAXMAT)                ! Simulation parameters.
→ COMMON/RSEED/ISEED1,ISEED2   ! Random number generator seeds.
```

As mentioned above, ILB(5) is an array of labels that describe the origin of secondary particles. It is assumed that the user has set ILB(1) equal to 1 (one) when a primary (source) particle history is initiated. Then, PENELOPE assigns the following labels to each particle in a shower;

ILB(1): generation of the particle. 1 for primary particles, 2 for their direct descendants, etc.

ILB(2): kind KPAR of the parent particle, only if ILB(1)>1 (secondary particles).

ILB(3): interaction mechanism ICOL (see above) that originated the particle, only when ILB(1)>1.

ILB(4): a non-zero value identifies particles emitted from atomic relaxation events and describes the atomic transition where the particle was released. The numerical value is $= Z \cdot 10^6 + \mathrm{IS1} \cdot 10^4 + \mathrm{IS2} \cdot 100 + \mathrm{IS3}$, where Z is the atomic number of the parent atom and IS1, IS3 and IS3 are the active atomic electron shells.

ILB(5): this label can be defined by the user; it is transferred to all descendants of the particle.

The ILB label values are delivered by subroutine SECPAR, through common TRACK, and remain unaltered during the simulation of the track.

Owing to the long execution time, the code will usually be run in batch mode. It is advisable to limit the simulation time rather than the number of tracks to be simulated, since the time required to follow each track is difficult to predict. To this end, one can link a clock routine to the simulation code and stop the computation after exhausting the allotted time; examples of clock routines for two different compilers are included in the PENELOPE distribution package.

6.1.3 Variance reduction

The subroutine package PENELOPE.F is intended to perform analogue simulation and, therefore, does not include any variance reduction methods. The source file PENVARED.F contains subroutines to perform splitting (VSPLIT), Russian roulette (VKILL) and interaction forcing (JUMPF, KNOCKF) in an automatic way. Splitting and Russian roulette (see section 1.6.2) do not require changes in PENELOPE; the necessary manipulations on the numbers and weights WGHT of particles could be done directly in the main program. Particles resulting from splitting are stored in the secondary stack by calling subroutine STORES.

Interaction forcing (section 1.6.1) implies changing the mean free paths of the forced interactions and, at the same time, redefining the weights of the generated secondary particles. In principle, it is possible to apply interaction forcing from the MAIN program by manipulating the interaction probabilities, that are made available through the named common block CJUMPO. These manipulations are performed automatically by calling the subroutines JUMPF and KNOCKF instead of JUMP and KNOCK.

Although these subroutines operate like "black boxes", they should be invoked with care. In general, it is advisable to prevent particle weights from reaching very large or very small values. In the first case, a very "heavy" particle can completely hide the information collected from many lighter particles. Conversely, it is not convenient to spend time simulating particles with very small weights, which contribute insignificant amounts to the scores. Notice also that repeated splitting and interaction forcing may easily lead to saturation of the secondary stack (the default stack size is 1000 particles). Hence, we usually apply interaction forcing only to primary particles.

6.2 Examples of MAIN programs

In general, the user must provide the MAIN program for each specific geometry. The distribution package of PENELOPE includes various examples of MAIN programs for simple geometries (slab and cylindrical) and for general quadric geometries with limited scoring.

 3620

In these examples, we assume that a single kind of particles is emitted from the radiation source. The programs can be easily generalized to the case of multi-particle sources with continuous (or discrete) energy spectra. For details on the operation of these codes, see section 6.2.4 below and the heading comments in the corresponding source files.

6.2.1 Program PENSLAB

The program PENSLAB simulates electron/photon showers within a material slab (see fig. 6.2). It illustrates the use of the simulation routines for the simplest geometry (as geometry operations are very simple, this program is faster than the ones described below). The slab is limited by the planes $z = 0$ and $z = t$, the thickness. The lateral extension of the slab is assumed to be infinite, i.e. much larger than the maximum range of the particles). Primary particles start with a given energy E_0 from a point source at a given "height" z_0 (positive or negative) on the z-axis, and moving in directions distributed uniformly in a spherical "sector" defined by its limiting polar angles, say θ_1 and θ_2, which is indicated by the hatched wedge in fig. 6.2. That is, to generate the initial direction, the polar cosine W $= \cos\theta$ is sampled uniformly in the interval from $\cos\theta_1$ to $\cos\theta_2$ and the azimuthal angle ϕ is sampled uniformly in $(0,2\pi)$. Thus, the case $\theta_1 = 0$ and $\theta_2 = 180$ deg corresponds to an isotropic source, whereas $\theta_1 = \theta_2 = 0$ defines a beam parallel to the z-axis. Notice that the complete arrangement has rotational invariance about the z-axis.

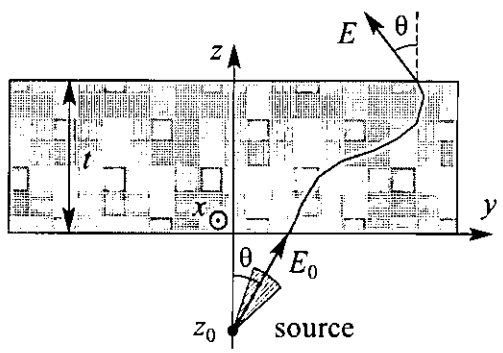


Figure 6.2: General planar geometry considered in PENSLAB.

PENSLAB generates detailed information on many quantities and distributions of physical interest. The output files contain a self-explanatory report of the simulation results, which consist of:

(i) Fractions of primary particles that are transmitted, backscattered and absorbed and a number of average quantities (track length within the sample; number of events of each kind per particle; energy, direction and lateral displacement of particles that leave the sample, etc.).

(ii) Energy distributions of transmitted and backscattered primary particles.

(iii) Angular distributions of transmitted and backscattered particles.

(iv) Depth-dose distribution (i.e. deposited energy per unit depth).

(v) Depth-distribution of deposited charge.

(vi) Distribution of energy deposited into the slab.

Each simulated continuous distribution is printed on a separate file (as a histogram), with a heading describing its content and in a format ready for visualization with a plotting program. The code computes and delivers the statistical uncertainties (3σ) of all evaluated quantities and distributions. Many authors quote these uncertainties as one standard deviation, which means that the probability for the actual value to lie outside the error bar is 0.317. We prefer to be more conservative and stay at the 3σ level, for which the probability of "missing the real value" is only 0.003.

The program PENSLAB and its predecessors have been intensively used during the last years to analyze the reliability of PENELOPE. They have been applied to a variety of experimental situations, covering a wide energy range. Benchmark comparisons with experimental data have been published elsewhere (Baró et al., 1995; Sempau et al., 1997).

WARNING: In the output files of PENSLAB (and also in those of the program PENCYL described below), the terms "transmitted" and "backscattered" are used to denote particles that leave the material system moving upwards (W > 0) and downwards (W < 0), respectively. Notice that this agrees with the usual meaning of these terms only when primary particles impinge on the system coming from below (i.e. with W > 0).

6.2.2 Program PENCYL

The program PENCYL simulates electron-photon showers in multilayered cylindrical structures. The material system consists of one or several layers of given thicknesses. Each layer contains a number of concentric homogeneous rings of given compositions and radii (and thickness equal to that of the layer). The layers are perpendicular to the z-axis and the centre of the rings in each layer is specified by giving its x and y coordinates. When all the centres are on the z-axis, the geometrical structure is symmetrical under rotations about the z-axis (see fig. 6.3).

Primary particles of a given kind, KPARP, are emitted from a point or extense (cylindrical) source, with its centre at the point (SX0, SY0, SZ0), either with fixed energy SE0 or with a specified (piecewise constant) energy spectrum. The initial direction of the primary particles is sampled uniformly inside a cone of (semi)aperture SALPHA and with central axis in the direction (STHETA, SPHI). Thus, the case SALPHA = 0 defines a monodirectional source and SALPHA = 180 deg corresponds to an isotropic source. When SX0 = SY0 = 0 and STHETA = 0 or 180 deg, the source is axially symmetrical about the z-axis.

In the distributed form of the program, we assume that both the source and the

material structure are symmetrical about the z-axis, because this eliminates the dependence on the azimuthal angle ϕ. The program takes advantage of this symmetry to tally 3D dose distributions. It is possible to consider geometries that are not axially symmetrical, but then the program only delivers values averaged over ϕ. To obtain the dependence of the angular distributions on the azimuthal angle, we need to increase the value of the parameter `NBPHM` (the maximum number of bins for ϕ, which is set equal to 1 in the distributed source file) and, in the input data file, set `NBPH` equal to `NBPHM`.

The source file `PENCYL.F` includes a (self-contained) set of geometry routines for tracking particles through multilayered cylindrical structures. These routines can be used for simulation even when the source is off-axis. Cylindrical geometries can be viewed with the program GVIEWC (which is similar to GVIEW2D and runs only under Microsoft Windows 9x). This program reads the geometry definition list from a file and displays a two-dimensional map of the materials intersected by the window plane. It is useful for debugging the geometry definition list.

`PENCYL` delivers detailed information on the transport and energy deposition, which includes energy and angular distributions of emerging particles, depth-dose distribution, depth-distribution of deposited charge, distributions of deposited energy in selected materials and 2D (depth-radius) dose and deposited charge distributions in selected bodies (cylinders). `PENCYL` can be directly used to study radiation transport in a wide variety of practical systems, e.g. planar ionization chambers, cylindrical scintillation detectors, solid-state detectors and multilayered structures.

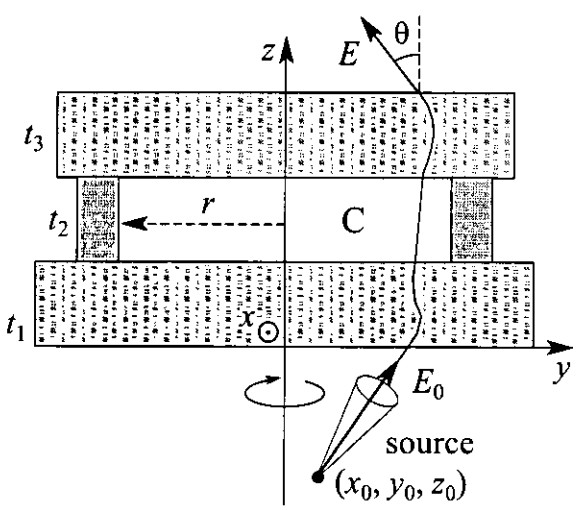


Figure 6.3: An example of cylindrical geometry, a cavity (C) with walls, with a point off-axis source. In this case, the material structure is symmetrical about the z-axis, but the radiation flux and other three-dimensional quantities (e.g. dose and deposited charge distributions) depend on the azimuthal angle ϕ.

6.2.3 Program PENDOSES

This MAIN program provides a practical example of simulation with complex material structures (quadric geometry only). It assumes a point source of primary particles at a given position $\mathbf{r}_0$ =(X0, Y0, Z0) which emits particles in directions uniformly distributed in a cone with (semi)aperture SALPHA and central axis in the direction (STHETA, SPHI) [the same direction distribution assumed in the code PENCYL]. The geometry of the material system is described by means of the package PENGEOM (chapter 5).

PENDOSES computes only the average energy deposited in each body per primary particle. With minor modifications, it also provides the probability distribution of the energy deposited in selected bodies or groups of bodies. It is a simple exercise to introduce a spatial grid, and the corresponding counters, and tally spatial dose distributions. Any future user of PENELOPE should become familiar with the programming details of PENDOSES before attempting her/his own application of PENELOPE.

6.2.4 Running the PENCYL program

The programs PENSLAB, PENCYL and PENDOSES operate in a similar way. They all read data from a corresponding input file and output the results in a number of files with fixed names[1]. The input files also have similar structures and formats. For concreteness, here we describe that of PENCYL, which is the most versatile of the example programs.

Each line in the input data file of PENCYL consists of a 6-character keyword (columns 1-6) followed either by numerical data (in free format) or by a character string, which start at the 8th column. Keywords are explicitly used/verified by the program (which is case sensitive!). Notice also that the order of the data lines is important. The keyword "------" (6 blanks, which we have denoted by "-") indicates comment lines, these can be placed anywhere after the end of the geometry definition list. The program ignores any text following the first blank after the last numerical datum, or after the character string, in each line (thus, in the table given below, the comments in square brackets are ignored by the program). Lines with certain keywords (e.g., "SPECTR") can appear an arbitrary number of times, limited only by the allocated amount of memory. The program assigns default values to many input variables; lines that declare default values may be removed from the input file. Notice that the default source is a pencil beam that moves upwards along the z-axis.

• Structure of the PENCYL input file (the 72-column rulers are just for visual aid, they do not form part of the input file)

```
....+....1....+....2....+....3....+....4....+....5....+....6....+....7..
TITLE  Title of the job, up to 65 characters.
GSTART >>>>>>>> Beginning of the geometry definition list.
```

[1]Warning: The programs overwrite older output files that are left in the working directory. You should save all result files on a separate directory before running the program again.

 24

```
LAYER  ZLOW,ZHIG                                    [Z_lower and Z_higher]
CENTRE XCEN,YCEN                                   [X_centre and Y_centre]
CYLIND M,RIN,ROUT                          [Material, R_inner and R_outer]
GEND   <<<<<<<< End of the geometry definition list.

       >>>>>>>> Source definition.
SKPAR  KPARP       [Primary particles: 1=electron, 2=photon, 3=positron]
SENERG SE0                 [Initial energy (monoenergetic sources only)]
SPECTR Ei,Pi                    [E bin: lower-end and total probability]
STHICK STHICK                                          [Source thickness]
SRADII SRIN,SROUT                          [Source inner and outer radii]
SPOSIT SX0,SY0,SZ0                    [Coordinates of the source centre]
SDIREC STHETA,SPHI                 [Beam axis direction angles, in deg]
SAPERT SALPHA                                   [Beam aperture, in deg]

       >>>>>>>> Material data and simulation parameters.
NMAT   NMAT                        [Number of different materials, .le.10]
SIMPAR M,EABS(1:3,M),C1,C2,WCC,WCR      [Sim. parameters for material M]
PFNAME filename_0.ext          [Material definition file, 18 characters]

       >>>>>>>> Counter array dimensions and pdf ranges.
NBE    NBE,EMIN,EMAX                [No. of energy bins, and E-interval]
NBTH   NBTH                       [No. of bins for the polar angle THETA]
NBPH   NBPH                    [No. of bins for the azimuthal angle PHI]
NBZ    NBZ                           [No. of bins for the Z-coordinate]
NBR    NBR                                          [No. of radial bins]
NBTL   NBTL,TLMIN,TLMAX       [No. of track-length bins and TL-interval]

       >>>>>>>> Additional distributions to be tallied.
ABSEN  MAT              [Tally the distr. of absorbed E in material MAT]
DOSE2D KL,KC          [Tally 2D dose and charge distribs. in body KL,KC]

       >>>>>>>> Interaction forcing.
IFORCE KL,KC,KPAR,ICOL,FORCE,WL,WH      [Interaction forcing parameters]

       >>>>>>>> Job properties
RESUME filename1.ext         [Resume from this dump file, 18 characters]
DUMPTO filename2.ext            [Generate this dump file, 18 characters]
NSIMSH NTOT           [Desired number of simulated showers, max=2**31-1]
RSEED  ISEED1,ISEED2             [Seeds of the random number generator]
TIME   ITIME                         [Allotted simulation time, in sec]
....+....1....+....2....+....3....+....4....+....5....+....6....+....7..
```

The following listing describes the function of each of the keywords, the accompanying

data and their default values.

TITLE ... title of the job (up to 65 characters).
Default: none (the input file must start with this line)

Geometry definition list ... begins with the line "GSTART" and ends with the line "GEND--" (notice the two blanks). The only allowed keywords in the geometry list are "GSTART", "LAYER-", "CENTRE", "CYLIND" and "GEND--". The line after "GSTART" must be a "LAYER-" line. Each "LAYER-" line contains the z-coordinates of its lower and upper limiting planes and is followed by a "CENTRE" line (optional) and by one or several "CYLIND" lines, which contain the inner and outer radii of the various concentric rings in the layer; empty layers are disregarded.

Layers must be defined in increasing order of heights, from bottom to top of the structure. If the "CENTRE" line is not entered, cylinders are assumed to be centered on the z-axis (XCEN = YCEN = 0.0). Cylinders have to be defined in increasing radial order, from the centre to the periphery. The two lengths in each "LAYER-" and "CYLIND" line must be entered in increasing order. The geometry definition list can be debugged/visualized with the code GVIEWC (operable only under Microsoft Windows 9x).

SKPAR ... kind of primary particle (1=electrons, 2=photons or 3=positrons).
Default: KPARP=1

SENERG ... for monoenergetic sources: initial energy SE0 of primary particles.
Default: SE0=1.0E6

SPECTR ... For sources with continuous (stepwise constant) energy spectra. Each "SPECTR" line gives the lower end-point of an energy bin of the source spectrum and the associated relative probability, integrated over the bin. Up to NSEM = 200 lines, in arbitrary order. The upper end of the spectrum is defined by entering a line with the upper energy value and null probability.
Default: none

STHICK ... thickness of the active volume of the source (cylinder).
Default: STHICK=0.0

SRADII ... inner and outer radii of the active source volume.
Defaults: SRIN=0.0, SROUT=0.0

SPOSIT ... coordinates of the centre of the source volume.
Defaults: SX0=SY0=0, SZ0=-1.0E15

SDIREC ... polar angle θ and azimuthal angle ϕ of the source beam axis direction, in deg.
Defaults: STHETA=0.0, SPHI=0.0

SAPERT ... angular aperture α of the source beam, in deg.
Default: SALPHA=0.0

NMAT ... number of different materials (up to 10 with the original program dimensions). Materials are identified by their ordering in PENELOPE's input material

 626

data file.
Default: NMAT=1

SIMPAR ... set of simulation parameters for the M-th material; absorption energies, EABS(1:3,M), elastic scattering parameters, C1(M) and C2(M), and cutoff energy losses for inelastic collisions and bremsstrahlung emission, WCC(M) and WCR(M). One line for each material.
Defaults: EABS(1,M)=EABS(3,M)=0.01*EPMAX, EABS(2,M)=0.001*EPMAX
C1(M)=C2(M)=0.1, WCC=EABS(1,M), WCR=EABS(2,M)
EPMAX is the maximum energy of all particles found in the simulation (depends on the source energies).

PFNAME ... name of PENELOPE's input material data file (18 characters).
Default: 'material.mat'

NBE ... number of energy bins and limits of the interval where energy distributions are tallied.
Defaults: NBE=100, EMIN=0.0, EMAX=EPMAX

NBTH ... number of bins for the polar angle θ.
Default: NBTH=90

NBPH ... number of bins for the azimuthal angle ϕ.
Default: NBPH=1 (azimuthal average).

NBZ ... number of bins for the z-coordinate.
Default: NBZ=100

NBR ... number of bins for the radial variable, $r = (x^2 + y^2)^{1/2}$.
Default: NBR=100

NBTL ... number of bins for track-length distributions of primary particles. Limits of the interval where these distributions are tallied.
Defaults: NBTL=100, TLMIN=0, TLMAX=2.5*RANGE(EPMAX,KPARP,1)

ABSEN ... indicates a material M for which we require the code to tally the distribution of absorbed energy (up to three different materials can be selected, a separate line for each).
Default: off

DOSE2D ... the program will tally 2D, depth-radius, dose and deposited charge distributions in the cylinder KC of layer KL. Up to three different cylinders (bodies) can be selected, a line for each body).
Default: off
Note: The labels KL, KC that identify a given cylinder are defined by the ordering in the input geometry list. These labels are written on the output geometry report.

IFORCE ... activates forcing of interactions of type ICOL of particles KPAR in cylinder KC of layer KL. FORCE is the forcing factor and WLOW, WHIG are the limits of the weight window where interaction forcing is active.
Default: no interaction forcing

RESUME ... the program will read the dump file named filename1.ext (18 characters) and resume the simulation from the point where it was left. Use this option very, *very* carefully. Make sure that the input data file is fully consistent with the one used to generate the dump file.
Default: off

DUMPTO ... generate a dump file named filename2.ext (18 characters) after completing the simulation run. This allows resuming the simulation to improve statistics.
Default: off

NSIMSH ... desired number of simulated showers. Notice that NTOT is an INTEGER*4 value and, hence, it cannot exceed $2^{31} - 1$.
Default: NTOT=2147 million

RSEED ... seeds of the random number generator.
Default: ISEED1=12345; ISEED2=54321

TIME ... allotted simulation time, in sec.
Default: ITIME=100

The program PENCYL (as well as the other example MAIN programs) is aborted when an incorrect input datum is found. The conflicting quantity usually appears in the last line of the output file. If the trouble is with arrays having dimensions smaller than required, the program indicates how the problem can be solved (this usually requires editing the source file, be careful).

The example of input file given below belongs to the PENCYL file set included in the distribution package. It corresponds to the simulation of a narrow photon beam with $E_0 = 1.25$ MeV (roughly the average energy of gamma rays from ^{60}Co) entering a 3"×3" NaI scintillation detector in an Al case, whose inner surface is partially covered by a layer of Al_2O_3, which diffuses scintillation light back to the crystal and the photomultiplier. In the material data file NAIAL.MAT, the order of the materials is NaI (MAT=1), Al_2O_3 (MAT=2) and Al (MAT=3). The incident beam photons move along the z-axis with $\theta = 0$ deg (i.e. upwards) and impinge normally on the surface of the detector. The geometry is shown schematically in the insets of fig. 6.4, which displays two of the distributions generated by PENCYL. The plotted distributions were obtained from 5 million random showers; the error bars represent statistical uncertainties (3σ), which are pretty small in this case.

• Example input file of the PENCYL code.

```
....+....1....+....2....+....3....+....4....+....5....+....6....+....7..
TITLE  Example from the distribution package...
GSTART NaI detector with Al cover and Al2O3 reflecting foil
LAYER       -0.24 -0.16  1
CENTRE       0.00  0.00
CYLIND   3   0.00  4.05
LAYER       -0.16  0.00  2
```

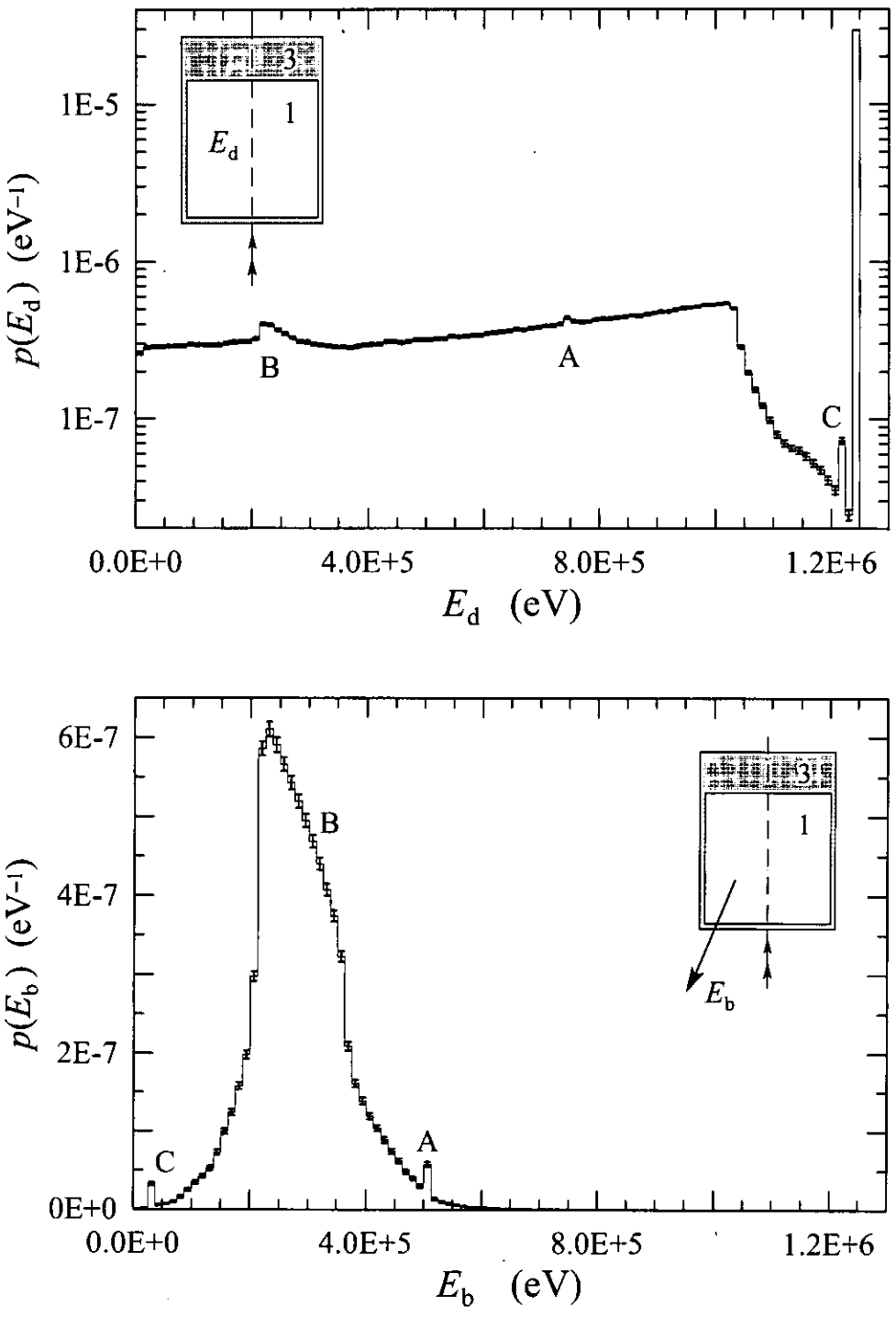


Figure 6.4: Partial results from `PENCYL` for the NaI photon detector described in the text. Top: distribution of energy deposited in the NaI crystal (`MAT=1`). Bottom: energy distribution of backscattered photons.

```
CYLIND   2   0.00  3.97
CYLIND   3   3.97  4.05
LAYER        0.00  7.72  3
CYLIND   1   0.00  3.81
CYLIND   2   3.81  3.97
CYLIND   3   3.97  4.05
LAYER        7.72  9.72  4
CYLIND   3   0.00  4.05
GEND
SKPAR  2         [Primary particles: 1=electron, 2=photon, 3=positron]
SENERG 1.25E6           [Initial energy (monoenergetic sources only)]
SPOSIT 0.0 0.0 -10.0                [Coordinates of the source centre]
NMAT   3                        [Number of different materials, .le.10]
SIMPAR 1 1.0E5 1000 1.0E5 0.1 0.1 1.0E4 1000   [M,EABS,C1,C2,WCC,WCR]
SIMPAR 2 1.0E5 1000 1.0E5 0.1 0.1 1.0E4 1000   [M.EABS,C1,C2,WCC,WCR]
SIMPAR 3 1.0E5 1000 1.0E5 0.1 0.1 1.0E4 1000   [M,EABS,C1,C2,WCC,WCR]
PFNAME naial.mat            [Material definition file, 18 characters]
ABSEN  1             [Tally the distr. of absorbed E in this material]
DOSE2D 3 1              [Tally 2D dose and charge distr. in this body]
RESUME lastdump.dat         [Read from this dump file, 18 characters]
DUMPTO lastdump.dat          [Generate this dump file, 18 characters]
NSIMSH 1000        [Desired number of simulated showers, max=2**31-1]
TIME   300                          [Allotted simulation time, in sec]
....+....1....+....2....+....3....+....4....+....5....+....6....+....7..
```

The upper plot in fig. 6.4 shows the distribution of energy E_d deposited into the NaI crystal volume (per primary photon). The lower figure displays the distribution (per primary photon) of the energy E_b of "backscattered" photons, i.e. photons that emerge from the system pointing "downwards", with $W = \cos\theta < 0$. These distributions show three conspicuous structures that arise from backscattering of incident photons in the crystal volume or in the Al backing (B), escape of one of the ~511 keV x-rays resulting from positron annihilation (A) and escape of ~30 keV iodine K x-rays (C). The peak A is so small because pair production is a relatively unlikely process for 1.25 MeV photons (the energy is too close to the threshold).

6.3 Selecting the simulation parameters

The speed and accuracy of the simulation of electrons and positrons is determined by the values of the simulation parameters E_{abs}, C_1, C_2, W_{cc}, W_{cr} and s_{max}, which are selected by the user for each material in the simulated structure[2]. Here we summarize the rules for assigning "safe" values to these parameters.

[2]To specify simulation parameters for a single body we can simply assign a specific material to this body, different from that of other bodies of the same composition.

The absorption energies $E_{\rm abs}$ are determined either by the characteristics of the experiment or by the required space resolution. If we want to tally dose or deposited-charge distributions, $E_{\rm abs}$ should be such that the residual range $R(E_{\rm abs})$ of electrons/positrons is less than the typical dimensions of the volume bins used to tally these distributions[3]. In other cases, it is advisable to run short simulations (for the considered body alone) with increasing values of $E_{\rm abs}$ (starting from 100 eV) to study the effect of this parameter on the results.

The allowed values of the elastic scattering parameters C_1 and C_2 are limited to the interval (0,0.2). For the present version 2001 of PENELOPE, these parameters have a very weak influence on the results, weaker than for previous versions of the code. As discussed in section 4.4.1, this is mostly due to the improved modelling of soft energy losses and to the consideration of the energy dependence of the hard mean free paths (see sections 4.2 and 4.3). Our recommended practice is to set $C_1 = C_2 = 0.05$, which is fairly conservative, as shown by the example given below. Before increasing the value of any of these parameters, it is advisable to perform short test simulations to verify that with the augmented parameter value the results remain essentially unaltered (and that the simulation runs faster; if there is no gain in speed, keep the conservative values).

We have already indicated that the cutoff energies $W_{\rm cc}$ and $W_{\rm cr}$ have a very weak influence on the accuracy of the results provided only that they are both smaller than the width of the bins used to tally energy distributions. When energy distributions are of no interest, our recommendation is setting these cutoff energies equal to one hundredth of the typical energy of primary particles.

The maximum allowed step length $s_{\rm max}$ (denoted by **DSMAX** in the FORTRAN source files) should be less than one tenth of the characteristic thickness of the body where the particle moves. This ensures that, on average, there will be more than 20 soft events[4] (hinges) along a typical electron/positron track within that body, which is enough to "wash out" the details of the artificial distributions used to sample these events. Notice however that PENELOPE internally forces the step length to be less than $\sim 3\lambda_{\rm T}^{(\rm h)}$ (see section 4.4). Therefore, for thick bodies (thicker than $\sim 20\lambda_{\rm T}^{(\rm h)}$), we can set $s_{\rm max} = 10^{35}$, or some other very large value, to switch off the external step-length control.

The **MAIN** program **PENSLAB** can be readily used to study the effect of the simulation parameters for a material body of a given characteristic thickness. As an example, figs. 6.5 and 6.6 display partial results from a **PENSLAB** simulation for a parallel electron beam of 500 keV impinging normally on the surface of a 200-μm-thick aluminium slab. The absorption energies were set equal to 10 keV (for all kinds of particles) and $W_{\rm cr}$ was given a negative value, which compels PENELOPE to set $W_{\rm cr} = 10$ eV and to disregard emission of soft bremsstrahlung (with $W < 10$ eV). We ran **PENSLAB** using $W_{\rm cc} = 0$ and $C_1 = C_2 = 0$; in this case, PENELOPE performs purely detailed, collision by

[3] PENELOPE prints tables of electron and positron ranges if subroutine **PEINIT** is invoked with **INFO=3** or larger.

[4] PENELOPE randomizes $s_{\rm max}$ in such a way that the actual step lengths never exceeds the value $s_{\rm max}$ set by the user and, on average, is equal to $s_{\rm max}/2$.

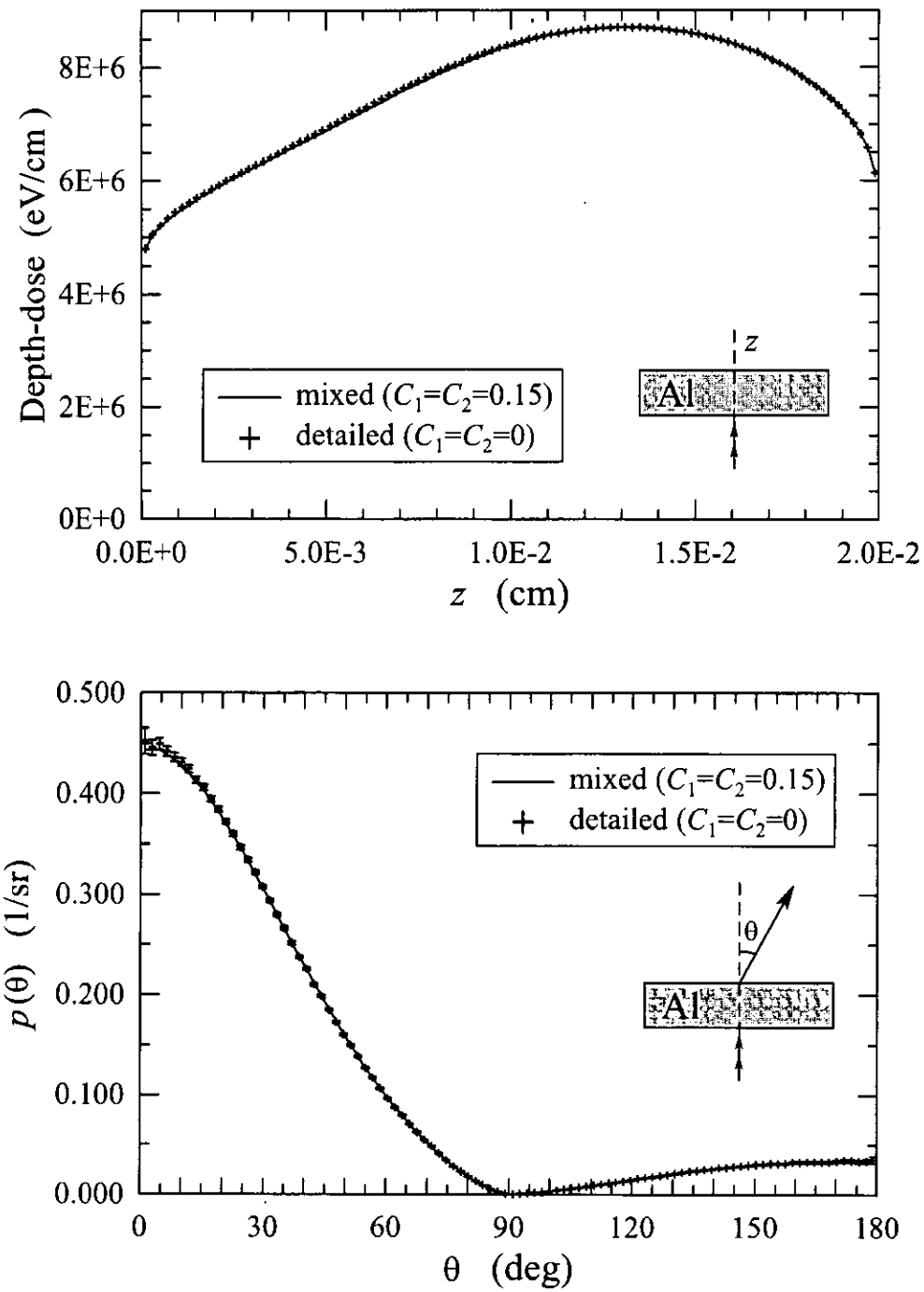


Figure 6.5: Results from **PENSLAB** for a 500 keV electron beam impinging normally on the surface of a 200 μm Al slab (further details are given in the text). Top: depth-dose distribution within the slab. Bottom: angular distribution of emerging (transmitted and backscattered) electrons (primary and secondary).

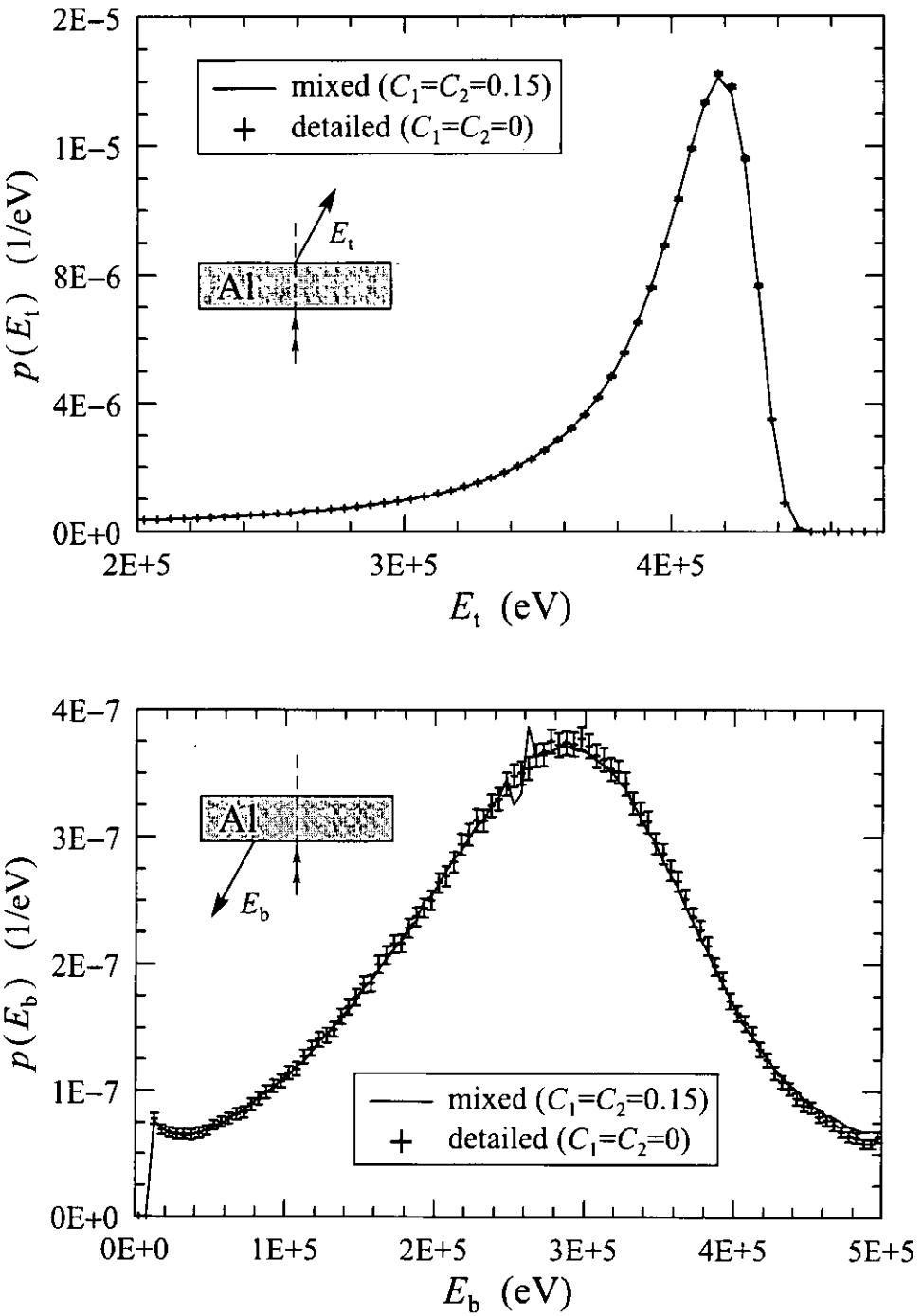


Figure 6.6: Results from PENSLAB for a 500 keV electron beam impinging normally on the surface of a 200 μm Al slab (further details are given in the text). Top: energy distribution of transmitted electrons. Bottom: energy distribution of backscattered electrons. Secondary electrons are included in both cases.

collision, simulation and, therefore, it provides exact results (affected only by statistical uncertainties and by inaccuracies of the physical interaction model). Differences between these results and those from mixed simulation are then completely attributable to the approximations in our mixed transport algorithm. To our knowledge, no other high-energy transport code allows detailed simulation and this kind of direct validation of the electron/positron transport mechanics.

In figs. 6.5 and 6.6 we compare results from the detailed simulation (7.5 million showers) with those from a mixed simulation using $W_{cc} = 1$ keV and $C_1 = C_2 = 0.15$ (20 million simulated showers); the error bars indicate statistical uncertainties (3σ). With these relatively high values of C_1 and C_2, mixed simulation is quite fast, the speed (generated showers per second) being about 45 times higher than that of the detailed simulation. As shown in the plots, mixed simulation results are practically equivalent to those from detailed simulation. It should be noted that backscattering, fig. 6.6b, is one of the most difficult cases to study, because it involves transport near and across an interface that is far from electronic equilibrium. The only visible artifact is a kind of singularity in the energy distribution of backscattered electrons at ~250 keV (which averages to the correct value and, therefore, would not be seen in a coarser energy grid). This artifact is also present in the energy distribution of transmitted electrons, but hardly visible in the scale of fig. 6.6a.

6.4 The code SHOWER

Monte Carlo simulation has proven to be a very valuable tool for education. In the past, radiation physics used to be considered as a tough subject, mostly because high-energy radiation is well outside the realm of daily experience. Nowadays, by simply running a transport simulation code on a personal computer we can learn more than from tens of obscure empirical formulas and numerical tables, and eventually "understand" many aspects of radiation transport (those for which we have run the simulation code and "digested" the results).

The PENELOPE distribution package includes a binary file named SHOWER that generates electron-photon showers within a slab (of one of the 279 materials defined in `PDCOMPOS.TAB`) and displays them (projected) on the computer screen plane. The current version operates only under Microsoft Windows 9x. The program is self-explanatory, and requires only a small amount of information from the user, which is entered from the keyboard, in response to prompts from the program. Electron, photon and positron tracks are displayed in different colors and intensities that vary with the energy of the particle. It is worth mentioning that the maximum number of showers that can be plotted in a single shot is limited to 50, because the screen may become too messy. Generating this small number of showers takes a short time, of the order of a few seconds, even on modest personal computers (provided only that the absorption energies are sensibly chosen).

Once on the graphical screen, the view plane can be rotated about the horizontal

screen axis by typing "r" and the rotation angle in degrees; the screen plane can also be rotated progressively, by 15 deg steps, by pressing the "enter" key repeatedly. Entering the single-character command "n" erases the screen and displays a new shower. Observation of single showers projected on a revolving plane gives a truly three-dimensional perspective of the transport process.

6.5 Installation

The FORTRAN77 source files of PENELOPE, the examples and auxiliary programs and the database are distributed as a single ZIP compressed file named `PENELOPE.ZIP`. To extract the files, keeping the directory structure, create the directory `PENELOPE` in your hard disk, copy the distribution file `PENELOPE.ZIP` into this directory and, from there, inflate (unzip) it. The directory structure and contents of the PENELOPE code system are the following:

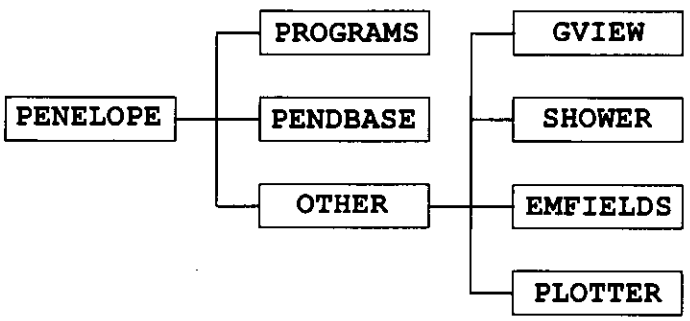


Figure 6.7: Directory tree of the PENELOPE code system.

- Subdirectory `PROGRAMS`. It contains the following 17 files:

  `MANUAL.TXT` ... abridged manual with general information.

  `PENELOPE.F` ... simulation subroutine package.

  `PENGEOM.F` ... modular quadric geometry subroutine package (handles systems with up to 250 surfaces and 125 bodies).

  `PENVARED.F` ... variance-reduction subroutines.

  `MATERIAL.F` ... main program to generate material data files.

  `PENSLAB.F` ... main program for particle transport in a slab.

  `PENSLAB.IN` ... sample input data file of `PENSLAB`.

  `AL.MAT` ... Material data file for `PENSLAB`.

  `PENCYL.F` ... main program for multilayered cylindrical geometries and axially symmetric beams.

PENCYL.IN ... sample input data file of PENCYL. Describes the same geometry as PENDOSES.GEO.

PENDOSES.F ... main program for arbitrary quadric geometries.

PENDOSES.IN ... sample input data file of PENDOSES.

PENDOSES.GEO ... geometry definition file for PENDOSES.

NAIAL.MAT ... material data file for PENCYL and PENDOSES. Illustrates the use of multiple materials.

TIMERMS.F ... clock subroutine, based on the subroutine GETTIM of Microsoft FORTRAN (version 5.0). It gives the execution time in seconds.

TIMERG77.F ... clock subroutine for the g77 FORTRAN compiler from the Free Software Foundation. The compact G77 for Win32 (Windows 9x/NT) package can be downloaded from
http://www.geocities.com/Athens/Olympus/5564
g77 is the default FORTRAN compiler in most Linux distributions.

TIMER1.F ... fake clock subroutine. To be used with "unfamiliar" compilers for which a proper timing routine is not known/available.

To get the executable file of MATERIAL, compile and link the files MATERIAL.F and PENELOPE.F. This executable file must be placed and run in the same subdirectory as the database files (PENDBASE).

The executable files of PENSLAB, PENCYL and PENDOSES are obtained by compiling and linking the following groups of source files (here we use the clock subroutine of the g77 compiler):

```
PENSLAB :  PENSLAB.F, PENELOPE.F, TIMERG77.F
PENCYL  :   PENCYL.F, PENELOPE.F, PENVARED.F, TIMERG77.F
PENDOSES: PENDOSES.F, PENELOPE.F,  PENGEOM.F, TIMERG77.F
```

The simulation programs are written in standard FORTRAN77 language, so that they should run on any computer. The only exception is the clock subroutine TIMERxxx.F, which must be adapted to your computer's compiler.

- Subdirectory PENDBASE. PENELOPE's database. 373 files with the extension ".TAB" and names beginning with the letters "PD" (for details, see section 6.1.1).

- Subdirectory OTHER. Consists of the following subdirectories,

GVIEW ... Contains the geometry viewers GVIEW2D, GVIEW3D and GVIEWC, that are operable only under Microsoft Windows 9x, and several examples of geometry definition files.

EMFIELDS ... Contains the subroutine package PENFIELD.F, which does simulation of electron/positron transport under external static magnetic and electric fields (see appendix C), and examples of programs that use it.

`SHOWER` ... Contains a single binary file named `SHOWER.EXE`, which operates only under Microsoft Windows 9x. This code generates electron-photon showers within a slab and displays them projected on the screen. To use the SHOWER viewer, just copy the file `SHOWER.EXE` into the directory `PENDBASE` and run it from there. This little tool is particularly useful for teaching purposes, it makes radiation physics "visible".

`PLOTTER` ... The programs `PENSLAB` and `PENCYL` generate multiple files with simulated probability distribution functions. Each output file has a heading describing its content, which is in a format ready for visualization with a plotting program. We use GNUPLOT, which is small in size, available for various platforms (including Linux and Microsoft Windows) and free (distribution sites are listed at the Gnuplot Central site, http://www.gnuplot.org). The directory `PLOTTER` contains GNUPLOT scripts that plot the probability distributions evaluated by the simulation codes on your terminal. For instance, after running `PENSLAB` you can visualize the results by simply 1) copying the file `PENSLAB.GNU` from the directory `PLOTTER` to the directory that contains the results and 2) entering the command "`GNUPLOT PENSLAB.GNU`" (or clicking the icon).

Appendix A

Collision kinematics

To cover the complete energy range of interest in radiation transport studies we use relativistic kinematics. Let $\tilde{P}$ denote the energy-momentum 4-vector of a particle, i.e.

$$\tilde{P} = (\mathcal{W}c^{-1}, \mathbf{p}), \tag{A.1}$$

where $\mathcal{W}$ and $\mathbf{p}$ are the total energy (including the rest energy) and momentum respectively and c is the velocity of light in vacuum. The product of 4-vectors, defined by

$$(\tilde{P}\tilde{P}') = \mathcal{W}\mathcal{W}'c^{-2} - \mathbf{p}\cdot\mathbf{p}', \tag{A.2}$$

is invariant under Lorentz transformations. The rest mass m of a particle determines the invariant length of its energy-momentum,

$$(\tilde{P}\tilde{P}) = \mathcal{W}^2c^{-2} - \mathbf{p}^2 = (mc)^2. \tag{A.3}$$

The kinetic energy E of a massive particle ($m \neq 0$) is defined as

$$E = \mathcal{W} - mc^2, \tag{A.4}$$

where mc^2 is the rest energy. The magnitude of the momentum is given by

$$(cp)^2 = E(E + 2mc^2). \tag{A.5}$$

In terms of the velocity $\mathbf{v}$ of the particle, we have

$$E = (\gamma - 1)mc^2 \quad \text{and} \quad \mathbf{p} = \beta\gamma mc\hat{\mathbf{v}}, \tag{A.6}$$

where

$$\beta \equiv \frac{v}{c} = \sqrt{\frac{\gamma^2 - 1}{\gamma^2}} = \sqrt{\frac{E(E + 2mc^2)}{(E + mc^2)^2}} \tag{A.7}$$

is the velocity of the particle in units of c and

$$\gamma \equiv \sqrt{\frac{1}{1 - \beta^2}} = \frac{E + mc^2}{mc^2} \tag{A.8}$$

is the total energy in units of the rest energy. From the relation (A.5), it follows that

$$E = \sqrt{(cp)^2 + m^2c^4} - mc^2 \tag{A.9}$$

and

$$\frac{\mathrm{d}p}{\mathrm{d}E} = \frac{1}{v} = \frac{1}{c\beta}. \tag{A.10}$$

For a photon (and any other particle with $m = 0$), the energy and momentum are related by

$$E = cp. \tag{A.11}$$

A.1 Two-body reactions

Consider a reaction in which a projectile "1" collides with a target "2" initially at rest in the laboratory frame of reference. We limit our study to the important case of two-body reactions in which the final products are two particles, "3" and "4". The kinematics of such reactions is governed by energy and momentum conservation.

We take the direction of movement of the projectile to be the z-axis, and set the x-axis in such a way that the reaction plane (i.e. the plane determined by the momenta of particles "1", "3" and "4") is the x-z plane. The energy-momentum 4-vectors of the projectile, the target and the reaction products are then (see fig. A.1)

$$\tilde{P}_1 = (\mathcal{W}_1 c^{-1}, 0, 0, p_1) \tag{A.12a}$$

$$\tilde{P}_2 = (m_2 c, 0, 0, 0) \tag{A.12b}$$

$$\tilde{P}_3 = (\mathcal{W}_3 c^{-1}, p_3 \sin\theta_3, 0, p_3 \cos\theta_3) \tag{A.12c}$$

$$\tilde{P}_4 = (\mathcal{W}_4 c^{-1}, -p_4 \sin\theta_4, 0, p_4 \cos\theta_4) \tag{A.12d}$$

Energy and momentum conservation is expressed by the 4-vector equation

$$\tilde{P}_1 + \tilde{P}_2 = \tilde{P}_3 + \tilde{P}_4. \tag{A.13}$$

From this equation, the angles of emergence of the final particles, θ_3 and θ_4, are uniquely determined by their energies, $\mathcal{W}_3$ and $\mathcal{W}_4$. Thus,

$$\begin{aligned} m_4^2c^2 &= (\tilde{P}_4\tilde{P}_4) = (\tilde{P}_1 + \tilde{P}_2 - \tilde{P}_3)(\tilde{P}_1 + \tilde{P}_2 - \tilde{P}_3) \\ &= (\tilde{P}_1\tilde{P}_1) + (\tilde{P}_2\tilde{P}_2) + (\tilde{P}_3\tilde{P}_3) + 2(\tilde{P}_1\tilde{P}_2) - 2(\tilde{P}_1\tilde{P}_3) - 2(\tilde{P}_2\tilde{P}_3) \\ &= m_1^2c^2 + m_2^2c^2 + m_3^2c^2 + 2\mathcal{W}_1\mathcal{W}_2c^{-2} \\ &\qquad - 2\left(\mathcal{W}_1\mathcal{W}_3c^{-2} - p_1p_3\cos\theta_3\right) - 2\mathcal{W}_2\mathcal{W}_3c^{-2}, \end{aligned} \tag{A.14}$$

and it follows that

$$\cos\theta_3 = \frac{m_4^2c^4 - m_1^2c^4 - m_2^2c^4 - m_3^2c^4 + 2\mathcal{W}_1(\mathcal{W}_3 - \mathcal{W}_2) + 2\mathcal{W}_2\mathcal{W}_3}{2\left(\mathcal{W}_1^2 - m_1^2c^4\right)^{1/2}\left(\mathcal{W}_3^2 - m_3^2c^4\right)^{1/2}}. \tag{A.15}$$

Clearly, by symmetry, we can obtain a corresponding expression for $\cos\theta_4$ by interchanging the indices 3 and 4

$$\cos\theta_4 = \frac{m_3^2c^4 - m_1^2c^4 - m_2^2c^4 - m_4^2c^4 + 2\mathcal{W}_1(\mathcal{W}_4 - \mathcal{W}_2) + 2\mathcal{W}_2\mathcal{W}_4}{2\,(\mathcal{W}_1^2 - m_1^2c^4)^{1/2}\,(\mathcal{W}_4^2 - m_4^2c^4)^{1/2}}. \tag{A.16}$$

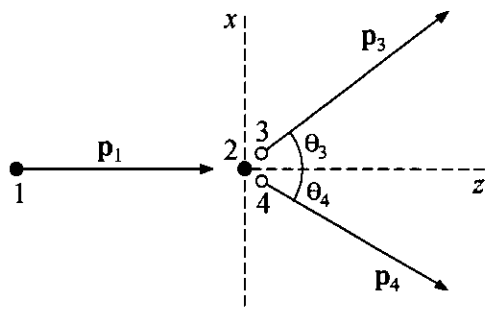


Figure A.1: Kinematics of two-body reactions.

The different two-body reactions found in Monte Carlo simulation of coupled electron-photon transport can be characterized by a single parameter, namely the energy of one of the particles that result from the reaction. The energy of the second particle is determined by energy conservation. Eqs. (A.15) and (A.16) then fix the polar angles, θ_3 and θ_4, of the final directions. Explicitly, we have

- Binary collisions of electrons and positrons with free electrons at rest.

Projectile:	Electron or positron	$m_1 = \mathrm{m_e}$,	$\mathcal{W}_1 = E + \mathrm{m_e}c^2$.
Target:	Electron	$m_2 = \mathrm{m_e}$,	$\mathcal{W}_2 = \mathrm{m_e}c^2$.
Scattered particle:		$m_3 = \mathrm{m_e}$,	$\mathcal{W}_3 = E - W + \mathrm{m_e}c^2$.
Recoil electron:		$m_4 = \mathrm{m_e}$,	$\mathcal{W}_4 = W + \mathrm{m_e}c^2$.

$$\cos\theta_3 = \left(\frac{E-W}{E}\,\frac{E+2\mathrm{m_e}c^2}{E-W+2\mathrm{m_e}c^2}\right)^{1/2}, \tag{A.17}$$

$$\cos\theta_4 = \left(\frac{W}{E}\,\frac{E+2\mathrm{m_e}c^2}{W+2\mathrm{m_e}c^2}\right)^{1/2}. \tag{A.18}$$

- Compton scattering of photons by free electrons at rest.

Projectile:	Photon	$m_1 = 0$,	$\mathcal{W}_1 = E \equiv \kappa\,\mathrm{m_e}c^2$.
Target:	Electron	$m_2 = \mathrm{m_e}$,	$\mathcal{W}_2 = \mathrm{m_e}c^2$.
Scattered photon:		$m_3 = 0$,	$\mathcal{W}_3 \equiv \tau E$.
Recoil electron:		$m_4 = \mathrm{m_e}$,	$\mathcal{W}_4 = \mathrm{m_e}c^2 + (1-\tau)E$.

$$\cos\theta_3 = \frac{1}{\kappa}\left(\kappa + 1 - \frac{1}{\tau}\right), \tag{A.19}$$

$$\cos\theta_4 = (\kappa+1)\left(\frac{1-\tau}{\kappa\,[2+\kappa(1-\tau)]}\right)^{1/2}. \tag{A.20}$$

- Annihilation of positrons with free electrons at rest.

Projectile:	Positron	$m_1 = \mathrm{m_e}$,	$\mathcal{W}_1 = E + \mathrm{m_e}c^2 \equiv \gamma\,\mathrm{m_e}c^2$.
Target:	Electron	$m_2 = \mathrm{m_e}$,	$\mathcal{W}_2 = \mathrm{m_e}c^2$.
Annihilation photons:		$m_3 = 0$,	$\mathcal{W}_3 \equiv \zeta(E + 2\mathrm{m_e}c^2)$.
		$m_4 = 0$,	$\mathcal{W}_4 = (1-\zeta)(E + 2\mathrm{m_e}c^2)$.

$$\cos\theta_3 = \left(\gamma^2 - 1\right)^{-1/2}(\gamma + 1 - 1/\zeta), \tag{A.21}$$

$$\cos\theta_4 = \left(\gamma^2 - 1\right)^{-1/2}\left(\gamma + 1 - \frac{1}{1-\zeta}\right). \tag{A.22}$$

A.1.1 Elastic scattering

By definition, elastic collisions keep the internal structure (i.e. the mass) of the projectile and target particles unaltered. Let us consider the kinematics of elastic collisions of a projectile of mass m ($= m_1 = m_3$) and kinetic energy E with a target particle of mass M ($= m_2 = m_4$) at rest (see fig. A.2). After the interaction, the target recoils with a certain kinetic energy W and the kinetic energy of the projectile is reduced to $E' = E - W$. The angular deflection of the projectile $\cos\theta$ and the energy transfer W are related through eq. (A.15), which now reads

$$\cos\theta = \frac{E(E+2mc^2) - W(E + mc^2 + Mc^2)}{\sqrt{E(E+2mc^2)\,(E-W)(E-W+2mc^2)}}. \tag{A.23}$$

The target recoil direction is given by eq. (A.16),

$$\cos\theta_{\mathrm{r}} = \frac{(E + mc^2 + Mc^2)W}{\sqrt{E(E+2mc^2)\,W(W+2mc^2)}}. \tag{A.24}$$

Solving eq. (A.23), we obtain the following expression for the energy transfer W corresponding to a given scattering angle θ,

$$W = \left[(E+mc^2)\sin^2\theta + Mc^2 - \cos\theta\sqrt{M^2c^4 - m^2c^4\sin^2\theta}\right] \times \frac{E(E+2mc^2)}{(E + mc^2 + Mc^2)^2 - E(E+2mc^2)\cos^2\theta}. \tag{A.25}$$

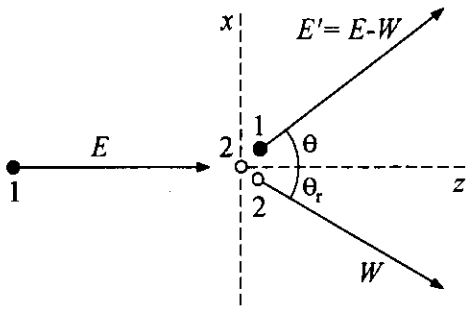


Figure A.2: Kinematics of elastic collisions.

In the case of collisions of particles with equal mass, $m = M$, this expression simplifies to

$$W = \frac{E(E + 2mc^2)\sin^2\theta}{E\sin^2\theta + 2mc^2} \qquad \text{if } M = m. \tag{A.26}$$

In this case, θ can only take values less than 90 deg. For $\theta = 90$ deg, we have $W = E$ (i.e. the full energy and momentum of the projectile are transferred to the target). Notice that for binary collisions of electrons and positrons ($m = \mathrm{m_e}$), the relation (A.26) becomes identical to (A.17).

For elastic collisions of electrons by atoms and ions, the mass of the target is much larger than that of the projectile and eq. (A.25) becomes

$$W = \frac{\left[(E + mc^2)\sin^2\theta + Mc^2(1 - \cos\theta)\right] E(E + 2mc^2)}{(E + Mc^2)^2 - E(E + 2mc^2)\cos^2\theta} \qquad \text{if } M \gg m. \tag{A.27}$$

The non-relativistic limit ($c \to \infty$) of this expression is

$$W = \frac{2m}{M}(1 - \cos\theta)E \qquad \text{if } M \gg m \text{ and } E \ll mc^2. \tag{A.28}$$

A.2 Inelastic collisions of charged particles

We consider here the kinematics of inelastic collisions of charged particles of mass m and velocity $\mathbf{v}$ as seen from a frame of reference where the stopping medium is at rest (laboratory frame). Let $\mathbf{p}$ and E be the momentum and the kinetic energy of the projectile just before an inelastic collision, the corresponding quantities after the collision are denoted by $\mathbf{p}'$ and $E' = E - W$, respectively. Evidently, for positrons the maximum energy loss is $W_{\max} = E$. In the case of ionization by electron impact, owing to the indistinguishability between the projectile and the ejected electron, the maximum energy loss is $W_{\max} \simeq E/2$ (see section 3.2). The momentum transfer in the collision is $\mathbf{q} \equiv \mathbf{p} - \mathbf{p}'$. It is customary to introduce the recoil energy Q defined by

$$Q(Q + 2\mathrm{m_e}c^2) = (cq)^2 = c^2\left(p^2 + p'^2 - 2pp'\cos\theta\right), \tag{A.29}$$

where m_e is the electron rest mass and $\theta = \arccos(\hat{\mathbf{p}}\cdot\hat{\mathbf{p}}')$ is the scattering angle. Equivalently, we can write

$$Q = \sqrt{(cq)^2 + m_e^2c^4} - m_ec^2. \tag{A.30}$$

Notice that, when the collision is with a free electron at rest, the energy loss is completely transformed into kinetic energy of the recoiling electron, i.e. $Q = W$. For collisions with bound electrons, the relation $Q \simeq W$ still holds for hard ionizing collisions (that is, when the energy transfer W is much larger than the ionization energy of the target electron so that binding effects are negligible).

The kinematically allowed recoil energies lie in the interval $Q_- < Q < Q_+$, with end points given by eq. (A.29) with $\cos\theta = +1$ and -1, respectively. That is

$$\begin{aligned} Q_\pm &= \sqrt{(cp \pm cp')^2 + m_e^2c^4} - m_ec^2 \\ &= \sqrt{\left[\sqrt{E(E+2mc^2)} \pm \sqrt{(E-W)(E-W+2mc^2)}\right]^2 + m_e^2c^4} - m_ec^2. \end{aligned} \tag{A.31}$$

Notice that, for $W < E$, Q_+ is larger than W and $Q_- < W$. When $W \ll E$, expression (A.31) is not suitable for evaluating Q_- since it involves the subtraction of two similar quantities. In this case, it is more convenient to use the approximate relation

$$cp - cp' \simeq c\left(-\frac{dp}{dE}W + \frac{1}{2}\frac{d^2p}{dE^2}W^2\right) = -\frac{W}{\beta}\left(1 + \frac{1}{2}\frac{1}{\gamma^2-1}\frac{W}{E+mc^2}\right) \tag{A.32}$$

and calculate Q_- as

$$Q_- \simeq \sqrt{(cp-cp')^2 + m_e^2c^4} - m_ec^2 \tag{A.33}$$

or, if $cp - cp' \ll m_ec^2$,

$$Q_- \simeq \frac{1}{2}\frac{(cp-cp')^2}{m_ec^2} - \frac{1}{8}\frac{(cp-cp')^4}{(m_ec^2)^3}. \tag{A.34}$$

Thus, for $E \gg W$,

$$Q_-(Q_- + 2m_ec^2) \simeq W^2/\beta^2. \tag{A.35}$$

In the non-relativistic limit,

$$Q \equiv q^2/2m_e, \qquad Q_\pm = \left[E^{1/2} \pm (E-W)^{1/2}\right]^2. \tag{A.36}$$

From (A.31), it is clear that the curves $Q = Q_-(W)$ and $Q = Q_+(W)$ vary monotonously with W and intersect at $W = E$. Thus, they define a single continuous function $W = W_m(Q)$ in the interval $0 < Q < Q_+(0)$. By solving the eqs. $Q = Q_\pm(W_m)$ we obtain

$$W_m(Q) = E + mc^2 - \sqrt{\left[\sqrt{E(E+2mc^2)} - \sqrt{Q(Q+2m_ec^2)}\right]^2 + m^2c^4}, \tag{A.37}$$

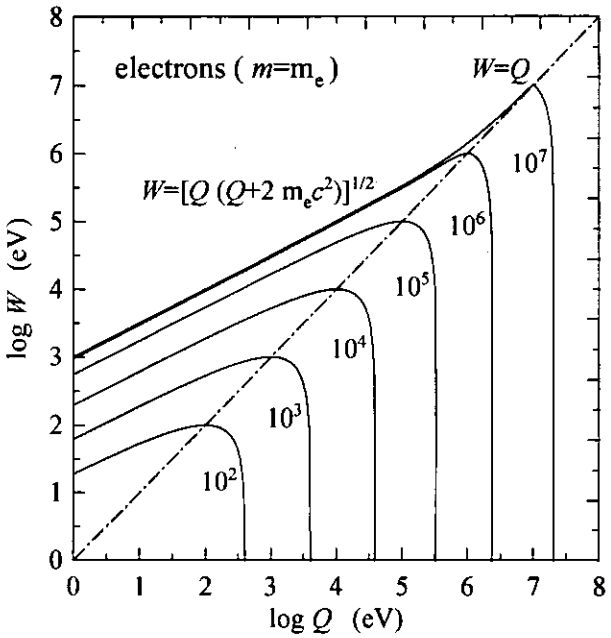


Figure A.3: Domains of kinematically allowed transitions in the (Q, W) plane for electrons/positrons. The curves represent the maximum allowed energy loss $W_{\rm m}(Q)$, given by eq. (A.37), for electrons with the indicated kinetic energies (in eV). When E increases, $W_{\rm m}(Q)$ approaches the vacuum photon line, $W = [Q(Q+2{\rm m_e}c^2)]^{1/2}$, which is an absolute upper bound for the allowed energy losses.

which, when $W \ll E$, reduces to

$$W_{\rm m}(Q) \simeq \beta\sqrt{Q(Q+2{\rm m_e}c^2)}. \tag{A.38}$$

Now it follows that, for given values of E and Q [$< Q_+(0)$], the only kinematically allowed values of the energy loss are those in the interval $0 < W < W_{\rm m}(Q)$ (see fig. A.3).

For a given energy loss W, the quantity

$$q_{\rm min} \equiv c^{-1}\sqrt{Q_-(Q_- + 2{\rm m_e}c^2)}, \tag{A.39}$$

is the minimum value of the momentum transfer in an inelastic collision, which occurs when $\theta = 0$. $q_{\rm min}$ is always larger than W/c. When the energy of the projectile increases, $\beta \to 1$ and $q_{\rm min}$ decreases approaching (but never reaching) the value W/c. It is worth recalling that a photon of energy W in vacuum has a linear momentum $q = W/c$ and, hence, interactions consisting of emission of bare photons would be located on the line $Q(Q + 2{\rm m_e}c^2) = W^2$ of the (Q,W) plane, the so-called vacuum photon line. This line, lies outside the kinematically allowed region, i.e. the "recoil" energy of the photon is less than Q_- (see fig. A.3). Therefore, when the target is a single atom, the emission of

photons by the projectile is not possible[1]. When the energy E of the projectile increases, Q_- decreases and tends to the photon line when β tends to unity. Hence, emission of photons by ultrarelativistic projectiles in low-density media is barely prevented by energy and momentum conservation. Generally speaking, as the interaction involves the exchange of a virtual photon, the DCS increases as the photon becomes more real, that is as we approach the photon line. For a dilute gas, this causes a gradual increase of the cross section with the projectile energy when $\beta \to 1$.

The scattering angle θ is related to the energy loss through

$$\cos\theta = \frac{(cp)^2 + (cp')^2 - Q(Q + 2\mathrm{m_e}c^2)}{2(cp)(cp)}. \tag{A.40}$$

The recoil angle θ_r between $\mathbf{p}$ and $\mathbf{q}$ is given by

$$\cos\theta_\mathrm{r} = \frac{(cp)^2 - (cp')^2 + (cq)^2}{2(cp)(cq)}, \tag{A.41}$$

which can also be written in the form

$$\cos^2\theta_\mathrm{r} = \frac{W^2/\beta^2}{Q(Q + 2\mathrm{m_e}c^2)}\left(1 + \frac{Q(Q + 2\mathrm{m_e}c^2) - W^2}{2W(E + mc^2)}\right)^2. \tag{A.42}$$

For heavy ($m \gg \mathrm{m_e}$) high-energy projectiles and collisions such that $Q \ll E$ and $W \ll E$,

$$\cos^2\theta_\mathrm{r} \simeq \frac{W^2/\beta^2}{Q(Q + 2\mathrm{m_e}c^2)} \simeq \frac{Q_-(Q_- + 2\mathrm{m_e}c^2)}{Q(Q + 2\mathrm{m_e}c^2)}. \tag{A.43}$$

[1]In a condensed medium, ultrarelativistic projectiles can emit real photons (Cerenkov radiation) under certain, quite restricting circumstances (see e.g. Jackson, 1975).

Appendix B

Numerical tools

B.1 Cubic spline interpolation

In this section we follow the presentation of Maron (1982). Suppose that a function $f(x)$ is given in numerical form, i.e. as a table of values

$$f_i = f(x_i) \qquad (i = 1, \ldots, N). \tag{B.1}$$

The points (knots) x_i do not need to be equispaced, but we assume that they are in (strictly) increasing order

$$x_1 < x_2 < \cdots < x_N. \tag{B.2}$$

A function $\varphi(x)$ is said to be an interpolating cubic spline if

1) It reduces to a cubic polynomial within each interval $[x_i, x_{i+1}]$, i.e. if $x_i \leq x \leq x_{i+1}$

$$\varphi(x) = a_i + b_i x + c_i x^2 + d_i x^3 \equiv p_i(x) \qquad (i = 1, \ldots, N-1). \tag{B.3}$$

2) The polynomial $p_i(x)$ matches the values of $f(x)$ at the endpoints of the i-th interval,

$$p_i(x_i) = f_i, \quad p_i(x_{i+1}) = f_{i+1} \qquad (i = 1, \ldots, N-1), \tag{B.4}$$

so that $\varphi(x)$ is continuous in $[x_1, x_N]$.

3) The first and second derivatives of $\varphi(x)$ are continuous in $[x_1, x_N]$

$$p_i'(x_{i+1}) = p_{i+1}'(x_{i+1}) \qquad (i = 1, \ldots, N-2), \tag{B.5}$$

$$p_i''(x_{i+1}) = p_{i+1}''(x_{i+1}) \qquad (i = 1, \ldots, N-2). \tag{B.6}$$

Consequently, the curve $y = \varphi(x)$ interpolates the table (B.1) and has a continuously turning tangent.

To obtain the spline coefficients a_i, b_i, c_i, d_i $(i = 1, \ldots, N-1)$ we start from the fact that $\varphi''(x)$ is linear in $[x_i, x_{i+1}]$. Introducing the quantities

$$h_i \equiv x_{i+1} - x_i \qquad (i = 1, \ldots, N-1) \tag{B.7}$$

and

$$\sigma_i \equiv \varphi''(x_i) \qquad (i = 1, \ldots, N), \tag{B.8}$$

we can write the obvious identity

$$p_i''(x) = \sigma_i \frac{x_{i+1} - x}{h_i} + \sigma_{i+1} \frac{x - x_i}{h_i} \qquad (i = 1, \ldots, N-1). \tag{B.9}$$

Notice that x_{i+1} must be larger than x_i in order to have $h_i > 0$. Integrating eq. (B.9) twice with respect to x, gives for $i = 1, \ldots, N-1$

$$p_i(x) = \sigma_i \frac{(x_{i+1} - x)^3}{6h_i} + \sigma_{i+1} \frac{(x - x_i)^3}{6h_i} + A_i(x - x_i) + B_i(x_{i+1} - x), \tag{B.10}$$

where A_i and B_i are constants. These can be determined by introducing the expression (B.10) into eqs. (B.4), this gives the pair of eqs.

$$\sigma_i \frac{h_i^2}{6} + B_i h_i = f_i \qquad \text{and} \qquad \sigma_{i+1} \frac{h_i^2}{6} + A_i h_i = f_{i+1}. \tag{B.11}$$

Finally, solving for A_i and B_i and substituting the result in (B.10), we obtain

$$\begin{aligned} p_i(x) = {} & \frac{\sigma_i}{6} \left[\frac{(x_{i+1} - x)^3}{h_i} - h_i(x_{i+1} - x) \right] + f_i \frac{x_{i+1} - x}{h_i} \\ & + \frac{\sigma_{i+1}}{6} \left[\frac{(x - x_i)^3}{h_i} - h_i(x - x_i) \right] + f_{i+1} \frac{x - x_i}{h_i}. \end{aligned} \tag{B.12}$$

To be able to use $\varphi(x)$ to approximate $f(x)$, we must find the second derivatives σ_i $(i = 1, \ldots, N)$. To this end, we impose the conditions (B.5). Differentiating (B.12) gives

$$p_i'(x) = \frac{\sigma_i}{6} \left[-\frac{3(x_{i+1} - x)^2}{h_i} + h_i \right] + \frac{\sigma_{i+1}}{6} \left[\frac{3(x - x_i)^2}{h_i} - h_i \right] + \delta_i, \tag{B.13}$$

where

$$\delta_i = \frac{y_{i+1} - y_i}{h_i}. \tag{B.14}$$

Hence,

$$p_i'(x_{i+1}) = \sigma_i \frac{h_i}{6} + \sigma_{i+1} \frac{h_i}{3} + \delta_i, \tag{B.15a}$$

$$p_i'(x_i) = -\sigma_i \frac{h_i}{3} - \sigma_{i+1} \frac{h_i}{6} + \delta_i \tag{B.15b}$$

and, similarly,

$$p'_{i+1}(x_{i+1}) = -\sigma_{i+1}\frac{h_{i+1}}{3} - \sigma_{i+2}\frac{h_{i+1}}{6} + \delta_{i+1}. \tag{B.15c}$$

Replacing (B.15a) and (B.15c) in (B.5), we obtain

$$h_i\sigma_i + 2(h_i + h_{i+1})\sigma_{i+1} + h_{i+1}\sigma_{i+2} = 6\,(\delta_{i+1} - \delta_i) \qquad (i = 1, \ldots, N-2). \tag{B.16}$$

The system (B.16) is linear in the N unknowns σ_i $(i = 1, \ldots, N)$. However, since it contains only $N-2$ equations, it is underdetermined. This means that we need either to add two additional (independent) equations or to fix arbitrarily two of the N unknowns. The usual practice is to adopt *endpoint strategies* that introduce constraints on the behaviour of $\varphi(x)$ near x_1 and x_N. An endpoint strategy fixes the values of σ_1 and σ_N, yielding an $(N-2)\times(N-2)$ system in the variables σ_i $(i = 2, \ldots, N-1)$. The resulting system is, in matrix form,

$$\begin{pmatrix} H_2 & h_2 & 0 & \cdots & 0 & 0 & 0 \\ h_2 & H_3 & h_3 & \cdots & 0 & 0 & 0 \\ 0 & h_3 & H_4 & \cdots & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & H_{N-3} & h_{N-3} & 0 \\ 0 & 0 & 0 & \cdots & h_{N-3} & H_{N-2} & h_{N-2} \\ 0 & 0 & 0 & \cdots & 0 & h_{N-2} & H_{N-1} \end{pmatrix} \begin{pmatrix} \sigma_2 \\ \sigma_3 \\ \sigma_4 \\ \vdots \\ \sigma_{N-3} \\ \sigma_{N-2} \\ \sigma_{N-1} \end{pmatrix} = \begin{pmatrix} D_2 \\ D_3 \\ D_4 \\ \vdots \\ D_{N-3} \\ D_{N-2} \\ D_{N-1} \end{pmatrix}, \tag{B.17}$$

where

$$H_i = 2(h_{i-1} + h_i) \qquad (i = 2, \ldots, N-1) \tag{B.18}$$

and

$$\begin{aligned} D_2 &= 6(\delta_2 - \delta_1) - h_1\sigma_1 \\ D_i &= 6(\delta_i - \delta_{i-1}) \qquad\qquad (i = 3, \ldots, N-2) \\ D_{N-1} &= 6(\delta_{N-1} - \delta_{N-2}) - h_{N-1}\sigma_N. \end{aligned} \tag{B.19}$$

(σ_1 and σ_N are removed from the first and last equations, respectively). The matrix of coefficients is symmetric, tridiagonal and diagonally dominant (the larger coefficients are in the diagonal), so that the system (B.17) can be easily (and accurately) solved by Gauss elimination. The spline coefficients a_i, b_i, c_i, d_i $(i = 1, \ldots, N-1)$ —see eq. (B.3)— can then be obtained by expanding the expressions (B.12):

$$\begin{aligned} a_i &= \frac{1}{6h_i}\left[\sigma_i x_{i+1}^3 - \sigma_{i+1}x_i^3 + 6\,(f_i x_{i+1} - f_{i+1}x_i)\right] + \frac{h_i}{6}\,(\sigma_{i+1}x_i - \sigma_i x_{i+1}), \\ b_i &= \frac{1}{2h_i}\left[\sigma_{i+1}x_i^2 - \sigma_i x_{i+1}^2 + 2\,(f_{i+1} - f_i)\right] + \frac{h_i}{6}\,(\sigma_i - \sigma_{i+1}), \\ c_i &= \frac{1}{2h_i}\,(\sigma_i x_{i+1} - \sigma_{i+1}x_i), \\ d_i &= \frac{1}{6h_i}\,(\sigma_{i+1} - \sigma_i). \end{aligned} \tag{B.20}$$

When accurate values of $f''(x)$ are known, the best strategy is to set $\sigma_1 = f''(x_1)$ and $\sigma_N = f''(x_N)$, since this will minimize the spline interpolation errors near the endpoints x_1 and x_N. Unfortunately, the exact values $f''(x_1)$ and $f''(x_N)$ are not always available.

The so-called *natural spline* corresponds to taking $\sigma_1 = \sigma_N = 0$. It results in a $y = \varphi(x)$ curve with the shape that would be taken by a flexible rod (such as a draughtman's spline) if it were bent around pegs at the knots but allowed to maintain its natural (straight) shape outside the interval $[x_1, x_N]$. Since $\sigma_1 = \sigma_N = 0$, extrapolation of $\varphi(x)$ outside the interval $[x_1, x_N]$ by straight segments gives a continuous function with continuous first and second derivatives [i.e. a cubic spline in $(-\infty, \infty)$].

The accuracy of the spline interpolation is mainly determined by the density of knots in the regions where $f(x)$ has strong variations. For constant, linear, quadratic and cubic functions the interpolation errors can be reduced to zero by using the exact values of σ_1 and σ_N (in these cases, however, the natural spline may introduce appreciable errors near the endpoints). It is important to keep in mind that a cubic polynomial has, at most, one inflexion point. As a consequence, we should have at least a knot between each pair of inflexion points of $f(x)$ to ensure proper interpolation. Special care must be taken when interpolating functions that have a practically constant value in a partial interval, since the spline tends to wiggle instead of staying constant. In this particular case, it may be more convenient to use linear interpolation.

Obviously, the interpolating cubic spline $\varphi(x)$ can be used not only to obtain interpolated values of $f(x)$ between the knots, but also to calculate integrals such as

$$\int_a^b f(x)\,\mathrm{d}x \simeq \int_a^b \varphi(x)\,\mathrm{d}x, \qquad x_1 \le a \quad \text{and} \quad b \le x_N, \tag{B.21}$$

analytically. It is worth noting that derivatives of $\varphi(x)$ other than the first one may differ significantly from those of $f(x)$.

To obtain the interpolated value $\varphi(x_c)$ —see eq. (B.3)— of $f(x)$ at the point x_c, we must first determine the interval $(x_i, x_{i+1}]$ that contains the point x_c. To reduce the effort to locate the point, we use the following binary search algorithm:

(i) Set $i = 1$ and $j = N$.

(ii) Set $k = [(i + j)/2]$.

(iii) If $x_k < x_c$, set $i = k$; otherwise set $j = k$.

(iv) If $j - i > 1$, go to step (ii).

(v) Deliver i.

Notice that the maximum delivered value of i is $N - 1$.

B.2 Numerical quadrature

In many cases, we need to calculate integrals of the form

$$\int_A^B f(z)\,\mathrm{d}z, \tag{B.22}$$

where the integrand is coded as an external function subprogram, which gives nominally exact values. These integrals are evaluated by using the `FORTRAN 77` external function `SUMGA`, which implements the twenty-point Gauss method with an adaptive bipartition scheme to allow for error control. This procedure is comparatively fast and is able to deal even with functions that have integrable singularities located at the endpoints of the interval $[A, B]$, a quite exceptional feature.

B.2.1 Gauss integration

We use the twenty-point Gauss formula (see e.g. Abramowitz and Stegun, 1974), given by

$$\int_a^b f(z)\,\mathrm{d}z = \frac{b-a}{2}\sum_{i=1}^{20} w_i f(z_i) \tag{B.23}$$

with

$$z_i = \frac{b-a}{2}x_i + \frac{b+a}{2}. \tag{B.24}$$

The abscissa x_i $(-1 < x_i < 1)$ is the i-th zero of the Legendre polynomial $P_{20}(x)$, the weights w_i are defined as

$$w_i = \frac{2}{(1-x_i^2)\,[P'_{20}(x_i)]^2}. \tag{B.25}$$

The numerical values of the abscissas and weights are given in table B.1. The difference between the exact value of the integral and the right-hand side of eq. (B.23) is

$$\Delta_{20} = \frac{(b-a)^{41}(20!)^4}{41\,(40!)^3} f^{(40)}(\xi), \tag{B.26}$$

where ξ is a point in the interval $[a, b]$.

The Gauss method gives an estimate of the integral of $f(z)$ over the interval $[a, b]$, which is obtained as a weighted sum of function values at fixed points inside the interval. We point out that (B.23) is an open formula, i.e. the value of the function at the endpoints of the interval is never required. Owing to this fact, function `SUMGA` can integrate functions that are singular at the endpoints. As an example, the integral of $f(x) = x^{-1/2}$ over the interval [0,1] is correctly evaluated. This would not be possible with a method based on a closed formula (i.e. one that uses the values of the integrand at the interval endpoints).

Table B.1: Abscissas and weights for twenty-point Gauss integration.

$\pm x_i$	w_i
7.6526521133497334D-02	1.5275338713072585D-01
2.2778585114164508D-01	1.4917298647260375D-01
3.7370608871541956D-01	1.4209610931838205D-01
5.1086700195082710D-01	1.3168863844917663D-01
6.3605368072651503D-01	1.1819453196151842D-01
7.4633190646015079D-01	1.0193011981724044D-01
8.3911697182221882D-01	8.3276741576704749D-02
9.1223442825132591D-01	6.2672048334109064D-02
9.6397192727791379D-01	4.0601429800386941D-02
9.9312859918509492D-01	1.7614007139152118D-02

B.2.2 Adaptive bipartition

Function SUMGA exploits the fact that the error Δ_{20}, eq. (B.26), of the calculated integral decreases when the interval length is reduced. Thus, halving the interval and applying the Gauss method to each of the two subintervals gives a much better estimate of the integral, provided only that the function $f(x)$ is smooth enough over the initial interval. Notice that the error decreases by a factor of about 2^{-40}(!).

The algorithm implemented in SUMGA is as follows. The integration interval (A, B) is successively halved so that each iteration gives a doubly finer partition of the initial interval. We use the term "n-subinterval" to denote the subintervals obtained in the n-th iteration. In each iteration, the integrals over the different n-subintervals are evaluated by the Gauss method, eq. (B.23). Consider that the integral over a given n-subinterval is S_1. In the following iteration, this n-subinterval is halved and the integrals over each of the two resulting $(n+1)$-subintervals are evaluated, giving values S_{1a} and S_{1b}. If $S_1' = S_{1a} + S_{1b}$ differs from S_1 in less than the selected tolerance, S_1' is the sought value of the integral in the considered n-subinterval; the value S_1' is then accumulated and this n-subinterval is no longer considered in subsequent iterations. Each iteration is likely to produce new holes (eliminated subintervals) in the regions where the function is smoother and, hence, the numerical effort progressively concentrates in the regions where $f(x)$ has stronger variations. The calculation terminates when the exploration of the interval (A, B) has been succesfully completed or when a clear indication of an anomalous behaviour of $f(x)$ is found (e.g. when there is a persistent increase of the number of remaining n-subintervals in each iteration). In the second case a warning message is printed in unit 6 and the control is returned to the calling program.

Appendix C

Electron/positron transport in electromagnetic fields

In this appendix, we consider the transport of electrons/positrons in static external electromagnetic (EM) fields, in vacuum and in condensed media. We assume that, in the region where particles move, there is an electric field $\boldsymbol{\mathcal{E}}$ and a magnetic field $\boldsymbol{\mathcal{B}}$, which are set up by external sources and do not vary with time. For practical purposes, we also consider that both $\boldsymbol{\mathcal{E}}$ and $\boldsymbol{\mathcal{B}}$ are continuous functions of the position vector $\mathbf{r}$.

The interactions with the medium will be described by means of PENELOPE. In each individual interaction event, the electron/positron looses a discrete amount of kinetic energy and changes its direction of motion. In the absence of EM fields, the electron travels freely between consecutive interaction events, i.e. following a straight trajectory segment at constant speed. To simulate electron transport with static external EM fields, we assume that the interaction properties of electrons with the medium are not substantially affected by the field. Consequently, to account for the effect of the EM field, we only need to consider that along each "free flight" the electron is driven by the EM force. With a proper selection of the simulation parameters (i.e. the energy loss and angular cutoff values), trajectory segments may have macroscopic lengths. Therefore, in material media it is appropriate to consider the macroscopic EM fields $\mathbf{D}$ and $\mathbf{H}$ rather than the microscopic fields $\boldsymbol{\mathcal{E}}$ and $\boldsymbol{\mathcal{B}}$.

It should be noted that, under the action of an electric field, the kinetic energy of the electron can vary substantially along a single trajectory segment. This conflicts with one of the basic assumptions in PENELOPE, namely that the energy of the particle stays practically constant along the segment. In practice, however, we can always limit the maximum segment length by means of the parameter s_{max}. Then, the effect of the EM field can be treated independently of that of the interactions with the medium. In other words, for simulation purposes, we only need an efficient method to generate particle trajectories in the EM field *in vacuum*. It is also important to recall that strong electric fields in material media accelerate unbound charged particles, even when they are at rest (i.e. electrons are never absorbed, simulated tracks can only terminate when

they leave the field volume). Injection of a single electron in the medium may give rise to a complex cascade of delta rays, that accelerate in the direction opposite to the electric field. To describe these cascades we need accurate cross sections for ionization by impact of low-energy electrons, much more accurate than the simple ones implemented in PENELOPE. Therefore, PENELOPE *is not* expected to yield a reliable description of this process. The simulation algorithm described here is applicable only to magnetic fields and; cautiously, to weak electric fields.

C.1 Tracking particles in vacuum.

Let us begin by describing a "brute force" method to calculate trajectories of charged particles in arbitrary static electric and magnetic fields in vacuum. We start from the Lorentz force equation[1] for an electron ($Z_0 = -1$) or positron ($Z_0 = +1$),

$$\frac{\mathrm{d}\mathbf{p}}{\mathrm{d}t} = Z_0 e(\boldsymbol{\mathcal{E}} + \mathbf{v}\times\boldsymbol{\mathcal{B}}), \tag{C.1}$$

which we write as

$$\frac{\mathrm{d}(\gamma\beta\hat{\mathbf{v}})}{\mathrm{d}t} = \frac{Z_0 e}{\mathrm{m_e}c}(\boldsymbol{\mathcal{E}} + c\beta\hat{\mathbf{v}}\times\boldsymbol{\mathcal{B}}), \tag{C.2}$$

with $\hat{\mathbf{v}} = \mathbf{v}/v$, $\beta = v/c$ and $\gamma = (1-\beta^2)^{-1/2}$. We note that

$$\frac{\mathrm{d}(\gamma\beta\hat{\mathbf{v}})}{\mathrm{d}t} = \gamma^3\frac{\mathrm{d}\beta}{\mathrm{d}t}\hat{\mathbf{v}} + \gamma\beta\frac{\mathrm{d}\hat{\mathbf{v}}}{\mathrm{d}t} \tag{C.3}$$

where the vectors $\hat{\mathbf{v}}$ and $\mathrm{d}\hat{\mathbf{v}}/\mathrm{d}t$ are orthogonal. Then, projecting eq. (C.2) into the directions of these two vectors, we obtain

$$\frac{\mathrm{d}\beta}{\mathrm{d}t} = \frac{Z_0 e}{\mathrm{m_e}c\gamma}(1-\beta^2)(\boldsymbol{\mathcal{E}}\cdot\hat{\mathbf{v}}) \tag{C.4}$$

and

$$\frac{\mathrm{d}\hat{\mathbf{v}}}{\mathrm{d}t} = \frac{Z_0 e}{\mathrm{m_e}c\beta\gamma}\left[\boldsymbol{\mathcal{E}} - (\boldsymbol{\mathcal{E}}\cdot\hat{\mathbf{v}})\hat{\mathbf{v}} + c\beta\hat{\mathbf{v}}\times\boldsymbol{\mathcal{B}}\right]. \tag{C.5}$$

It then follows that

$$\begin{aligned}\frac{\mathrm{d}\beta\hat{\mathbf{v}}}{\mathrm{d}t} &= \frac{\mathrm{d}\beta}{\mathrm{d}t}\hat{\mathbf{v}} + \beta\frac{\mathrm{d}\hat{\mathbf{v}}}{\mathrm{d}t} \\ &= \frac{Z_0 e}{\mathrm{m_e}c\gamma}\left[\boldsymbol{\mathcal{E}} - \beta^2(\boldsymbol{\mathcal{E}}\cdot\hat{\mathbf{v}})\hat{\mathbf{v}} + c\beta\hat{\mathbf{v}}\times\boldsymbol{\mathcal{B}}\right],\end{aligned} \tag{C.6}$$

which we cast in the form

$$\frac{\mathrm{d}\mathbf{v}}{\mathrm{d}t} = \mathbf{A}, \qquad \mathbf{A} \equiv \frac{Z_0 e}{\mathrm{m_e}\gamma}\left[\boldsymbol{\mathcal{E}} - \beta^2(\boldsymbol{\mathcal{E}}\cdot\hat{\mathbf{v}})\hat{\mathbf{v}} + c\beta\hat{\mathbf{v}}\times\boldsymbol{\mathcal{B}}\right]. \tag{C.7}$$

[1] In this appendix, electromagnetic quantities are expressed in SI units.

 8653

Notice that, for arbitrary fields $\boldsymbol{\mathcal{E}}$ and $\boldsymbol{\mathcal{B}}$, the "acceleration" $\mathbf{A}$ is a function of the particle's position $\mathbf{r}$, energy E and direction of motion $\hat{\mathbf{v}}$.

Implicit integration of eq. (C.7) gives the equations of motion

$$\mathbf{v}(t) = \mathbf{v}_0 + \int_0^t \mathbf{A}(\mathbf{r}(t'), E(t'), \hat{\mathbf{v}}(t'))\,\mathrm{d}t', \tag{C.8}$$

$$\mathbf{r}(t) = \mathbf{r}_0 + t\,\mathbf{v}_0 + \int_0^t \mathbf{v}(t')\,\mathrm{d}t'. \tag{C.9}$$

Evidently, these equations are too complex for straight application in a simulation code and we must have recourse to approximate solution methods. We shall adopt the approach proposed by Bielajew (1988), which is well suited to transport simulations. The basic idea is to split the trajectory into a number of conveniently short steps such that *the acceleration* $\mathbf{A}$ *does not change much over the course of a step.* Along each step, we then have

$$\mathbf{v}(t) = \mathbf{v}_0 + t\,\mathbf{A}(\mathbf{r}_0, E_0, \hat{\mathbf{v}}_0) \tag{C.10}$$

$$\mathbf{r}(t) = \mathbf{r}_0 + t\,\mathbf{v}_0 + t^2\,\frac{1}{2}\,\mathbf{A}(\mathbf{r}_0, E_0, \hat{\mathbf{v}}_0), \tag{C.11}$$

where the subscript "0" indicates values of the various quantities at the starting point ($t = 0$). The traveled path length s and the flying time t are related by

$$t = \int_0^s \frac{\mathrm{d}s'}{v}, \tag{C.12}$$

which, to first order becomes

$$t = \frac{s}{v_0}\left(1 - \frac{v(t) - v_0}{2v(t)}\right). \tag{C.13}$$

Then, to first order in the electromagnetic force,

$$\mathbf{v}(s) = \mathbf{v}_0 + s\,\frac{\mathbf{A}(\mathbf{r}_0, E_0, \hat{\mathbf{v}}_0)}{c\beta_0}$$

$$\mathbf{r}(s) = \mathbf{r}_0 + s\,\hat{\mathbf{v}}_0 + s^2\,\frac{1}{2}\,\frac{\mathbf{A}(\mathbf{r}_0, E_0, \hat{\mathbf{v}}_0)}{c^2\beta_0^2}.$$

That is,

$$\mathbf{r}(s) = \mathbf{r}_0 + s\,\hat{\mathbf{v}}_0 + s^2\,\frac{1}{2}\,\frac{Z_0 e\,[\boldsymbol{\mathcal{E}}_0 - \beta_0^2(\boldsymbol{\mathcal{E}}_0\cdot\hat{\mathbf{v}}_0)\hat{\mathbf{v}}_0 + c\beta_0\hat{\mathbf{v}}_0\times\boldsymbol{\mathcal{B}}_0]}{\mathrm{m_e}c^2\,\gamma_0\beta_0^2}. \tag{C.14}$$

The particle's velocity can be calculated directly from eq. (C.10), which to first order gives

$$\mathbf{v}(s) = \mathbf{v}_0 + \Delta\mathbf{v} \tag{C.15}$$

with

$$\Delta\mathbf{v} = s\,\frac{Z_0 e\,[\boldsymbol{\mathcal{E}}_0 - \beta_0^2(\boldsymbol{\mathcal{E}}_0\cdot\hat{\mathbf{v}}_0)\hat{\mathbf{v}}_0 + c\beta_0\hat{\mathbf{v}}_0\times\boldsymbol{\mathcal{B}}_0]}{\mathrm{m_e}c\,\gamma_0\beta_0}. \tag{C.16}$$

In the tracking algorithm, the velocity is used to determine the direction vector at the end of the step,

$$\hat{\mathbf{v}}(s) = \frac{\mathbf{v}_0 + \Delta\mathbf{v}}{|\mathbf{v}_0 + \Delta\mathbf{v}|}. \tag{C.17}$$

Owing to the action of the electromagnetic force, the kinetic energy E of the particle varies along the step. As the trajectory is accurate only to first order, it is not advisable to compute the kinetic energy from the velocity of the particle. It is preferable to calculate $E(t)$ as

$$E(s) = E_0 + Z_0 e\,[\varphi(\mathbf{r}_0) - \varphi(\mathbf{r}(s))] \tag{C.18}$$

where $\varphi(\mathbf{r})$ is the electrostatic potential, $\boldsymbol{\mathcal{E}} = -\nabla\varphi$. Notice that this ensures energy conservation, i.e. it gives the exact energy variation in going from the initial to the final position.

This tracking method is valid only if
1) the fields do not change too much along the step

$$\frac{|\boldsymbol{\mathcal{E}}(\mathbf{r}(s)) - \boldsymbol{\mathcal{E}}(\mathbf{r}_0)|}{|\boldsymbol{\mathcal{E}}(\mathbf{r}_0)|} < \delta_{\boldsymbol{\mathcal{E}}} \ll 1, \qquad \frac{|\boldsymbol{\mathcal{B}}(\mathbf{r}(s)) - \boldsymbol{\mathcal{B}}(\mathbf{r}_0)|}{|\boldsymbol{\mathcal{B}}(\mathbf{r}_0)|} < \delta_{\boldsymbol{\mathcal{B}}} \ll 1 \tag{C.19}$$

and
2) the relative changes in kinetic energy and velocity (or direction of motion) are small

$$\left|\frac{E(s) - E_0}{E_0}\right| < \delta_E \ll 1, \qquad \frac{|\Delta\mathbf{v}|}{v_0} < \delta_{\mathrm{v}} \ll 1. \tag{C.20}$$

These conditions set an upper limit on the allowed step length, $s_{\max}$, which depends on the local fields *and* on the energy and direction of the particle. The method is robust, in the sense that it converges to the exact trajectory when the maximum allowed step length tends to zero. In practical calculations, we shall specify the values of the δ-parameters (which should be of the order of 0.05 or less) and consider step lengths consistent with the above conditions. Thus, the smallness of the δ-parameters determines the accuracy of the generated trajectories.

To test the accuracy of a tracking algorithm, it is useful to consider the special cases of a uniform electric field (with $\boldsymbol{\mathcal{B}} = 0$) and a uniform magnetic field (with $\boldsymbol{\mathcal{E}} = 0$), which admit relatively simple analytical solutions of the equations of motion.

C.1.1 Uniform electric fields

Let us study first the case of a uniform electric field $\boldsymbol{\mathcal{E}}$. The equation of the trajectory of an electron/positron that starts at $t = 0$ from the point $\mathbf{r}_0$ with velocity $\mathbf{v}_0$ can be expressed in the form (adapted from Bielajew, 1988)

$$\mathbf{r}(t) = \mathbf{r}_0 + t\mathbf{v}_{0\perp} + \frac{1}{a}\left[\cosh(act) - 1 + \frac{v_{0\parallel}}{c}\sinh(act)\right]\hat{\boldsymbol{\mathcal{E}}}, \tag{C.21}$$

 655

where $\mathbf{v}_{0\|}$ and $\mathbf{v}_{0\perp}$ are the components of $\mathbf{v}_0$ parallel and perpendicular to the direction of the field,

$$\mathbf{v}_{0\|} = (\mathbf{v}_0 \cdot \hat{\boldsymbol{\mathcal{E}}})\hat{\boldsymbol{\mathcal{E}}}, \qquad \mathbf{v}_{0\perp} = \mathbf{v}_0 - (\mathbf{v}_0 \cdot \hat{\boldsymbol{\mathcal{E}}})\hat{\boldsymbol{\mathcal{E}}} \tag{C.22}$$

and

$$a \equiv \frac{Z_0 e \mathcal{E}}{\mathrm{m}_e c^2 \gamma_0} = \frac{Z_0 e \mathcal{E}}{E_0}. \tag{C.23}$$

The velocity of the particle is

$$\begin{aligned} \mathbf{v}(t) &= \mathbf{v}_{0\perp} + \left[c \sinh(act) + v_{0\|} \cosh(act)\right] \hat{\boldsymbol{\mathcal{E}}} \\ &= \mathbf{v}_0 + \left\{c \sinh(act) + v_{0\|} [\cosh(act) - 1]\right\} \hat{\boldsymbol{\mathcal{E}}}. \end{aligned} \tag{C.24}$$

Since the scalar potential for the constant field is $\varphi(\mathbf{r}) = -\boldsymbol{\mathcal{E}} \cdot \mathbf{r}$, the kinetic energy of the particle varies with time and is given by

$$E(t) = E_0 - Z_0 e \boldsymbol{\mathcal{E}} \cdot [\mathbf{r}_0 - \mathbf{r}(t)]. \tag{C.25}$$

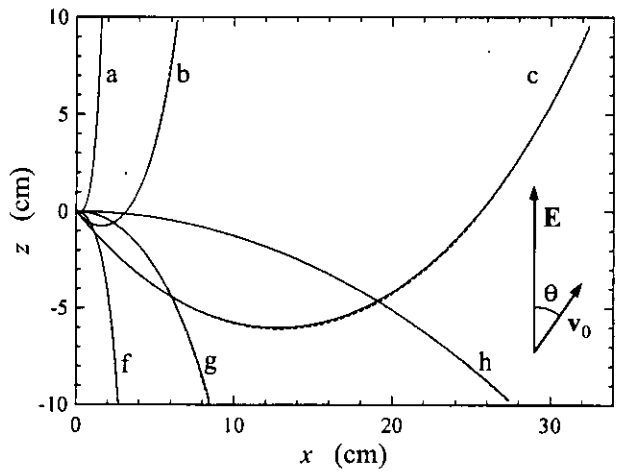


Figure C.1: Trajectories of electrons and positrons in a uniform electric field of 511 kV/cm. Continuous curves represent exact trajectories obtained from eq. (C.21). The dashed lines are obtained by using the first-order numerical tracking method described by eqs. (C.14)-(C.20) with $\delta_{\boldsymbol{\mathcal{E}}} = \delta_E = \delta_{\mathbf{v}} = 0.02$. The displayed trajectories correspond to the following cases. a: positrons, $E_0 = 0.1$ MeV, $\theta = 135$ deg. b: positrons, $E_0 = 1$ MeV, $\theta = 135$ deg. c: positrons, $E_0 = 10$ MeV, $\theta = 135$ deg. f: electrons, $E_0 = 0.1$ MeV, $\theta = 90$ deg. g: electrons, $E_0 = 1$ MeV, $\theta = 90$ deg. h: electrons, $E_0 = 10$ MeV, $\theta = 90$ deg.

Fig. C.1 displays trajectories of electrons and positrons with various initial energies and directions of motion in a uniform electric field of 511 kV/cm directed along the positive z-axis. Particles start from the origin ($\mathbf{r}_0 = 0$), with initial velocity in the xz-plane forming an angle θ with the field, i.e. $\mathbf{v}_0 = (\sin\theta, 0, \cos\theta)$, so that the whole

 3656

trajectories lie in the xz-plane. Continuous curves represent exact trajectories obtained from the analytical formula (C.21). The dashed curves are the results from the first-order tracking algorithm described above [eqs. (C.14)-(C.20)] with $\delta_{\mathcal{E}} = \delta_E = \delta_\nu = 0.02$. We show three positron trajectories with initial energies of 0.1, 1 and 10 MeV, initially moving in the direction $\theta = 135$ deg. Three trajectories of electrons that initially move perpendicularly to the field ($\theta = 90$ deg) with energies of 0.1, 1 and 10 MeV are also depicted. We see that the tracking algorithm gives quite accurate results. The error can be further reduced, if required, by using shorter steps, i.e. smaller δ-values.

C.1.2 Uniform magnetic fields

We now consider the motion of an electron/positron, with initial position $\mathbf{r}_0$ and velocity $\mathbf{v}_0$, in a uniform magnetic field $\boldsymbol{\mathcal{B}}$. Since the magnetic force is perpendicular to the velocity, the field does not alter the energy of the particle and the speed $v(t) = v_0$ is a constant of the motion. It is convenient to introduce the precession frequency vector $\boldsymbol{\omega}$, defined by (notice the sign)

$$\boldsymbol{\omega} \equiv -\frac{Z_0 e \boldsymbol{\mathcal{B}}}{\mathrm{m_e}\gamma} = -\frac{Z_0 e c^2 \boldsymbol{\mathcal{B}}}{E_0}, \tag{C.26}$$

and split the velocity $\mathbf{v}$ into its components parallel and perpendicular to $\boldsymbol{\omega}$,

$$\mathbf{v}_{\parallel} = (\mathbf{v}\cdot\hat{\boldsymbol{\omega}})\hat{\boldsymbol{\omega}}, \qquad \mathbf{v}_{\perp} = \mathbf{v} - (\mathbf{v}\cdot\hat{\boldsymbol{\omega}})\hat{\boldsymbol{\omega}}. \tag{C.27}$$

Then, the equation of motion (C.7) becomes

$$\frac{\mathrm{d}\mathbf{v}_{\parallel}}{\mathrm{d}t} = 0, \qquad \frac{\mathrm{d}\mathbf{v}_{\perp}}{\mathrm{d}t} = \boldsymbol{\omega}\times\mathbf{v}_{\perp}. \tag{C.28}$$

The first of these eqs. says that the particle moves with constant velocity $\mathbf{v}_{0\parallel}$ along the direction of the magnetic field. From the second eq. we see that, in the plane perpendicular to $\boldsymbol{\mathcal{B}}$, the particle describes a circle with angular frequency ω and speed $v_{0\perp}$ (which is a constant of the motion). The radius of the circle is $R_\perp = v_{0\perp}/\omega$. That is, the trajectory is an helix with central axis along the $\boldsymbol{\mathcal{B}}$ direction, radius $R_\perp$ and pitch angle $\alpha = \arctan(v_{0\parallel}/v_{0\perp})$. The helix is right-handed for electrons and left-handed for positrons (see fig. C.2).

In terms of the path length $s = tv_0$, the equation of motion takes the form

$$\mathbf{r}(s) = \mathbf{r}_0 + \frac{s}{v_0}\mathbf{v}_{0\parallel} + R_\perp\left[\cos(s_\perp/R_\perp) - 1\right](\hat{\mathbf{v}}_{0\perp}\times\hat{\boldsymbol{\omega}}) + R_\perp\sin(s_\perp/R_\perp)\hat{\mathbf{v}}_{0\perp}, \tag{C.29}$$

where $\hat{\mathbf{v}}_{0\perp} \equiv \mathbf{v}_{0\perp}/v_{0\perp}$ and $s_\perp = sv_{0\perp}/v_0$. Equivalently,

$$\mathbf{r}(s) = \mathbf{r}_0 + s\hat{\mathbf{v}}_0 - \frac{s}{v_0}\mathbf{v}_{0\perp} + \frac{1}{\omega}\left[\cos(s\omega/v_0) - 1\right](\mathbf{v}_{0\perp}\times\hat{\boldsymbol{\omega}}) + \frac{1}{\omega}\sin(s\omega/v_0)\mathbf{v}_{0\perp}. \tag{C.30}$$

After the path length s, the particle velocity is

$$\mathbf{v}(s) = v_0\frac{\mathrm{d}\mathbf{r}}{\mathrm{d}s} = \mathbf{v}_0 + \left[\cos(s\omega/v_0) - 1\right]\mathbf{v}_{0\perp} - \sin(s\omega/v_0)(\mathbf{v}_{0\perp}\times\hat{\boldsymbol{\omega}}). \tag{C.31}$$

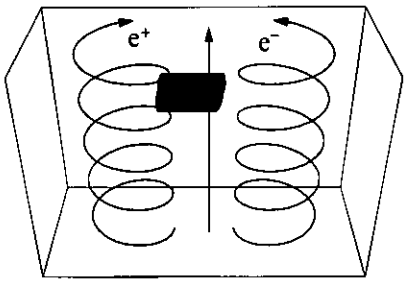


Figure C.2: Trajectories of electrons and positrons in a uniform magnetic field. The two particles start from the base plane with equal initial velocities.

In fig. C.3 we compare exact trajectories of electrons and positrons in a uniform magnetic field obtained from the analytical formula (C.30) with results from the first-order tracking algorithm [eqs. (C.14)-(C.20)] with $\delta_{\mathcal{B}} = \delta_E = \delta_v = 0.02$. The field strength is 0.2 tesla. The depicted trajectories correspond to 0.5 MeV electrons (a) and 3 MeV positrons (b) that initially move in a direction forming an angle of 45 deg with the field. We see that the numerical algorithm is quite accurate for small path lengths, but it deteriorates rapidly for increasing s. In principle, the accuracy of the algorithm can be improved by reducing the value of δ_v, i.e. the length of the step length. In practice, however, this is not convenient because it implies a considerable increase of numerical work, which can be easily avoided.

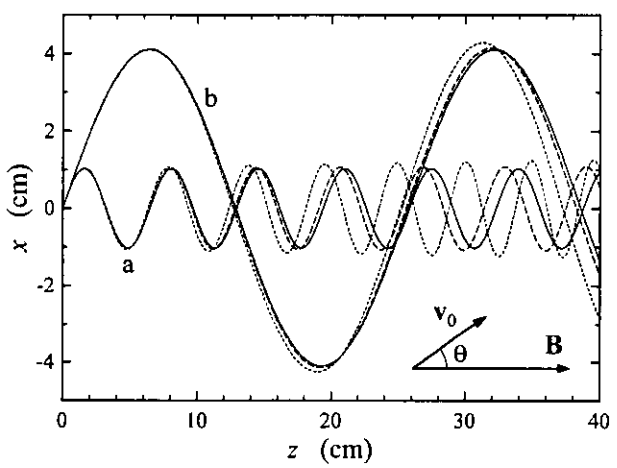


Figure C.3: Trajectories of electrons and positrons in a uniform magnetic field of 0.2 tesla. Continuous curves are exact trajectories calculated from eq. (C.30). The short-dashed lines are obtained by using the numerical tracking method described in the text with $\delta_v = 0.02$. Long-dashed curves are the results from the tracking algorithm with $\delta_v = 0.005$. a: electrons, $E_0 = 0.5$ MeV, $\theta = 45$ deg. b: positrons, $E_0 = 3$ MeV, $\theta = 45$ deg.

C.2 Exact tracking in homogeneous magnetic fields

In our first-order tracking algorithm [see eqs. (C.14) and (C.16)], the effects of the electric and magnetic fields are decoupled, i.e. they can be evaluated separately. For uniform electric fields, the algorithm offers a satisfactory solution since it usually admits relatively large step lengths. In the case of uniform magnetic fields (with $\boldsymbol{\mathcal{E}} = 0$), the kinetic energy is a constant of the motion and the only effective constraint on the step length is that the change in direction $|\Delta\mathbf{v}|/v_0$ has to be small. Since the particle trajectories on the plane perpendicular to the field $\boldsymbol{\mathcal{B}}$ are circles and the first-order algorithm generates each step as a parabolic segment, we need to move in sub-steps of length much less than the radius $R_\perp$ (i.e. δ_v must be given a very small value) and this makes the calculation slow. On the other hand, the action of the uniform magnetic field is described by simple analytical expressions [eqs. (C.30) and (C.31)], that are amenable for direct use in the simulation code. These arguments suggest the following obvious modification of the tracking algorithm.

As before, we assume that the fields are essentially constant along each trajectory step and write

$$\mathbf{r}(s) = \mathbf{r}_0 + s\hat{\mathbf{v}}_0 + (\Delta\mathbf{r})_{\boldsymbol{\mathcal{E}}} + (\Delta\mathbf{r})_{\boldsymbol{\mathcal{B}}}, \tag{C.32}$$

where $(\Delta\mathbf{r})_{\boldsymbol{\mathcal{E}}}$ and $(\Delta\mathbf{r})_{\boldsymbol{\mathcal{B}}}$ are the displacements caused by the electric and magnetic fields, respectively. For $(\Delta\mathbf{r})_{\boldsymbol{\mathcal{E}}}$ we use the first-order approximation [see eq. (C.14)],

$$(\Delta\mathbf{r})_{\boldsymbol{\mathcal{E}}} = s^2\,\frac{1}{2}\,\frac{Z_0 e\,[\boldsymbol{\mathcal{E}}_0 - \beta_0^2(\boldsymbol{\mathcal{E}}_0\cdot\hat{\mathbf{v}}_0)\hat{\mathbf{v}}_0]}{\mathrm{m_e}c^2\,\gamma_0\beta_0^2}. \tag{C.33}$$

The displacement caused by the magnetic field is evaluated using the result (C.30), i.e.

$$(\Delta\mathbf{r})_{\boldsymbol{\mathcal{B}}} = -\frac{s}{v_0}\mathbf{v}_{0\perp} + \frac{1}{\omega}\,[\cos(s\omega/v_0) - 1]\,(\mathbf{v}_{0\perp}\times\hat{\boldsymbol{\omega}}) + \frac{1}{\omega}\sin(s\omega/v_0)\mathbf{v}_{0\perp} \tag{C.34}$$

with

$$\boldsymbol{\omega} \equiv -\frac{Z_0 e c^2 \boldsymbol{\mathcal{B}}_0}{E_0}, \qquad \text{and} \qquad \mathbf{v}_{0\perp} = \mathbf{v}_0 - (\mathbf{v}_0\cdot\hat{\boldsymbol{\omega}})\hat{\boldsymbol{\omega}}. \tag{C.35}$$

Similarly, the particle velocity along the step is expressed as

$$\mathbf{v}(s) = \mathbf{v}_0 + (\Delta\mathbf{v})_{\boldsymbol{\mathcal{E}}} + (\Delta\mathbf{v})_{\boldsymbol{\mathcal{B}}} \tag{C.36}$$

with [see eqs. (C.16) and (C.31)]

$$(\Delta\mathbf{v})_{\boldsymbol{\mathcal{E}}} = s\,\frac{Z_0 e\,[\boldsymbol{\mathcal{E}}_0 - \beta_0^2(\boldsymbol{\mathcal{E}}_0\cdot\hat{\mathbf{v}}_0)\hat{\mathbf{v}}_0]}{\mathrm{m_e}c\,\gamma_0\beta_0} \tag{C.37}$$

and

$$(\Delta\mathbf{v})_{\boldsymbol{\mathcal{B}}} = [\cos(s\omega/v_0) - 1]\,\mathbf{v}_{0\perp} - \sin(s\omega/v_0)(\mathbf{v}_{0\perp}\times\hat{\boldsymbol{\omega}}). \tag{C.38}$$

In our implementation of this tracking algorithm, the allowed step lengths s are limited by the following constraints [see eqs. (C.19) and (C.20)]

$$\frac{|\boldsymbol{\mathcal{E}}(\mathbf{r}(s)) - \boldsymbol{\mathcal{E}}(\mathbf{r}_0)|}{|\boldsymbol{\mathcal{E}}(\mathbf{r}_0)|} < \delta_{\boldsymbol{\mathcal{E}}} \ll 1, \qquad \frac{|\boldsymbol{\mathcal{B}}(\mathbf{r}(s)) - \boldsymbol{\mathcal{B}}(\mathbf{r}_0)|}{|\boldsymbol{\mathcal{B}}(\mathbf{r}_0)|} < \delta_{\boldsymbol{\mathcal{B}}} \ll 1 \tag{C.39}$$

and

$$\left|\frac{E(s) - E_0}{E_0}\right| < \delta_E \ll 1, \qquad \frac{\left|(\Delta\mathbf{v})_{\boldsymbol{\mathcal{E}}} + (\Delta\mathbf{v})_{\boldsymbol{\mathcal{B}}}\right|}{v_0} < \delta_{\mathrm{v}} \ll 1. \tag{C.40}$$

The algorithm is robust, i.e. the accuracy of the generated trajectories increases when the δ-parameters are reduced. In many practical cases, a good compromise between accuracy and simulation speed is obtained by setting $\delta_{\boldsymbol{\mathcal{E}}} = \delta_{\boldsymbol{\mathcal{B}}} = \delta_E = \delta_{\mathrm{v}} = 0.02$. Notice that, in the case of a uniform magnetic field, the tracking algorithm is now exact, irrespective of the step length.

This tracking algorithm has been implemented in the subroutine package PENFIELD, which is devised to work linked to PENELOPE and PENGEOM. To simulate radiation transport in a given field/material configuration, the user must provide the steering main program as well as specific routines that define the EM field (see the examples and comments in the source file `PENFIELD.F`).

Bibliography

Acosta E., X. Llovet, E. Coleoni, J.A. Riveros and F. Salvat (1998), "Monte Carlo simulation of x-ray emission by kilovolt electron bombardment", *J. Appl. Phys.* **83**, 6038–6049.

Abramowitz M. and I.A. Stegun (1974), eds., *Handbook of Mathematical Functions* (Dover, New York).

Andreo P. (1991), "Monte Carlo techniques in medical radiation physics", *Phys. Med. Biol.* **36**, 861–920.

Andreo P. and A. Brahme (1984), "Restricted energy-loss straggling and multiple scattering of electrons in mixed Monte Carlo procedures", *Rad. Res.* **100**, 16–29.

Baró J., M. Roteta, J.M. Fernández-Varea and F. Salvat (1994a), "Analytical cross sections for Monte Carlo simulation of photon transport", *Radiat. Phys. Chem.* **44**, 531–552.

Baró J., J. Sempau, J.M. Fernández-Varea and F. Salvat (1994b), "Simplified Monte Carlo simulation of elastic electron scattering in limited media", *Nucl. Instrum. Meth. B* **84**, 465–483.

Baró J., J. Sempau, J.M. Fernández-Varea and F. Salvat (1995), "PENELOPE: an algorithm for Monte Carlo simulation of the penetration and energy loss of electrons and positrons in matter", *Nucl. Instrum. Meth. B* **100**, 31–46.

Benedito E., J.M. Fernández-Varea and F. Salvat (2001), "Mixed simulation of the multiple elastic scattering of electrons and positrons using partial-wave differential cross sections", *Nucl. Instrum. Meth. B* **174**, 91–110.

Berger M.J. (1963), "Monte Carlo calculation of the penetration and diffusion of fast charged particles", in *Methods in Computational Physics*, vol. 1, eds. B. Alder, S. Fernbach and M. Rotenberg (Academic Press, New York) pp. 135–215.

Berger M.J. (1992), "ESTAR, PSTAR and ASTAR: computer programs for calculating stopping-power and range tables for electrons, protons and helium ions", *Report NISTIR 4999* (National Institute of Standards and Technology, Gaithersburg, MD).

Berger M.J. (1998), "Applicability of the condensed-random-walk Monte Carlo method at low energies in high-Z materials", *Radiat. Phys. Chem.* **53**, 191–203.

Berger M.J. and J.H. Hubbell (1987), "XCOM: Photon Cross Sections on a Personal Computer", *Report NBSIR 87-3597* (National Bureau of Standards, Gaithersburg, MD).

Berger M.J. and S.M. Seltzer (1972), "Response functions for sodium iodide scintillation detectors", *Nucl. Instrum. Meth.* **104**, 317–332.

Berger M.J. and S.M. Seltzer (1982), "Stopping Power of Electrons and Positrons", *Report NBSIR 82-2550* (National Bureau of Standards, Gaithersburg, MD).

Berger M.J. and S.M. Seltzer (1988), chapters 7, 8 and 9, in *Monte Carlo Transport of Electrons and Photons*, eds. T.M. Jenkins, W.R. Nelson and A. Rindi (Plenum, New York).

Bethe, H.A. (1930), "Zur Theorie des Durchgangs schneller Korpurkularstrahlen durch Materie", *Ann. Physik* **5**, 325–400.

Bethe, H.A. (1932), "Bremsformel für Elektronen relativistischer Geschwindigkeit", *Z. Physik* **76**, 293–299.

Bethe H.A. and W. Heitler (1934), "On the stopping of fast particles and on the creation of positive electrons", *Proc. R. Soc. (London) A* **146**, 83–112.

Bhabha H.J. (1936), "The scattering of positrons by electrons with exchange on Dirac's theory of the positron", *Proc. R. Soc. (London) A* **154**, 195–206.

Bielajew A.F. (1988), "Electron transport in $\vec{E}$ and $\vec{B}$ fields", in *Monte Carlo Transport of Electrons and Photons*, eds. T.M. Jenkins, W.R. Nelson and A. Rindi (Plenum, New York) pp. 421–434.

Bielajew A.F. (1995), "HOWFAR and HOWNEAR: Geometry Modeling for Monte Carlo Particle Transport", *Report PIRS-0341* (National Research Council of Canada, Ottawa).

Bielajew A.F. and D.W.O. Rogers (1987), "PRESTA: The parameter reduced electron-step transport algorithm for electron Monte Carlo transport", *Nucl. Instrum. Meth. B* **18**, 165–181.

Bielajew A.F. and D.W.O. Rogers (1988), "Variance-reduction techniques", in *Monte Carlo Transport of Electrons and Photons*, eds. T.M. Jenkins, W.R. Nelson and A. Rindi (Plenum, New York) pp. 407–419.

Bielajew A.F. and F. Salvat (2001), "Improved electron transport mechanics in the PENELOPE Monte-Carlo model", *Nucl. Instrum. Meth. B* **173**, 332–343.

Biggs F., L.B. Mendelsohn and J.B. Mann (1975), "Hartree-Fock Compton profiles for the elements", *At. Data Nucl. Data Tables* **16**, 201–309.

Blunck O. and S. Leisegang (1950), "Zum Energieverlust schneller Elektronen in dünnen Schichten", *Z. Physik* **128**, 500–505.

Born M. (1969), *Atomic Physics* (Blackie and Son, London).

Bransden B.H. and C.J. Joachain (1983), *Physics of Atoms and Molecules* (Longman, Essex, England).

Briesmeister J.F. (1997), "MCNP—A general Monte Carlo N-particle transport code", *Report LA-12625-M Version 4B* (Los Alamos National Laboratory, Los Alamos, NM).

Brun R., F. Bruyant, M. Maire, A.C. McPherson and P. Zanarini (1986), "GEANT3", *Report DD/EE/84-1* (CERN, Geneva).

Brusa D., G. Stutz, J.A. Riveros, J.M. Fernández-Varea and F. Salvat (1996), "Fast sampling algorithm for the simulation of photon Compton scattering", *Nucl. Instrum. Meth. A* **379**, 167–175.

Chan H.-P. and K. Doi (1983), "The validity of Monte Carlo simulation in studies of scattered radiation in diagnostic radiology", *Phys. Med. Biol.* **28**, 109–129.

Cooper M. (1971), "Compton scattering and electron momentum distributions", *Adv. Phys.* **20**, 453–491.

Cullen D.E., M.H. Chen, J.H. Hubbell, S.T. Perkins, E.F. Plechaty, J.A. Rathkopf and J.H. Scofield (1989), "Tables and graphs of photon-interaction cross sections from 10 eV to 100 GeV derived from the LLNL evaluated photon data library (EPDL)", *Report UCRL-50400* vol. 6, rev. 4, parts A and B. (Lawrence Livermore National Laboratory, Livermore, CA).

Cullen D.E., J.H. Hubbell and L. Kissel (1997), "EPDL97 The evaluated data library, '97 version", *Report UCRL-50400* vol. 6, rev. 5 (Lawrence Livermore National Laboratory, Livermore, CA).

Davies H., H.A. Bethe and L.C. Maximon (1954), "Theory of bremsstrahlung and pair production. II. Integral cross section for pair production" *Phys. Rev.* **93**, 788–795.

Desclaux J.P. (1975), "A multiconfiguration relativistic Dirac-Fock program", *Comput. Phys. Commun.* **9**, 31–45. Erratum: ibid. **13** (1977) 71.

Doyle P.A. and P.S. Turner (1968), "Relativistic Hartree-Fock X-ray and electron scattering factors", *Acta Cryst. A* **24**, 390–397.

Edmonds A.R. (1960), *Angular Momentum in Quantum Mechanics*, 2nd edition (Princeton University Press, Princeton, NJ).

Fano U. (1954), "Inelastic collisions and the Molière theory of multiple scattering", *Phys. Rev.* **93**, 117–120.

Fano U. (1963), "Penetration of protons, alpha particles and mesons", *Ann. Rev. Nucl. Sci.* **13**, 1–66.

Fernández-Varea J.M., R. Mayol, D. Liljequist and F. Salvat (1993a) "Inelastic scattering of electrons in solids from a generalized oscillator strength model using optical and photoelectric data", *J. Phys: Condens. Matter* **5**, 3593–3610.

Fernández-Varea J.M., R. Mayol, J. Baró and F. Salvat (1993b), "On the theory and simulation of multiple elastic scattering of electrons", *Nucl. Instrum. Meth. B* **73**, 447–473.

Goudsmit S. and J.L. Saunderson (1940a), "Multiple scattering of electrons", *Phys. Rev.* **57**, 24–29.

Goudsmit S. and J.L. Saunderson (1940b), "Multiple scattering of electrons. II", *Phys. Rev.* **58**, 36–42.

Halbleib J.A., R.P. Kensek, T.A. Mehlhorn, G.D. Valdez, S.M. Seltzer and M.J. Berger (1992), "ITS version 3.0: the integrated TIGER series of coupled electron/photon Monte Carlo transport codes", *Report SAND91-1634* (Sandia National Laboratories, Albuquerque, NM).

Haug E. (1975), "Bremsstrahlung and pair production in the field of free electrons", *Z. Naturforsch.* **30a**, 1099–1113.

Hayward E. and J. Hubbell (1954), "The albedo of various materials for 1-Mev photons", *Phys. Rev.* **93**, 955–956.

Heinrich K.F.J. and D.E. Newbury (1991), eds., *Electron Probe Quantitation* (Plenum Press, New York).

Heitler W. (1954), *The Quantum Theory of Radiation* (Oxford Univ. Press, London).

Hubbell J.H. (1989), "Bibliography and current status of K, L, and higher shells fluorescence yields for computations of photon energy-absorption coefficients", *Report NISTIR 89-4144* (National Institute of Standards and Technology, Gaithersburg, MD).

Hubbell J.H., H.A. Gimm and I. Øverbø (1980), "Pair, triplet, and total atomic cross sections (and mass attenuation coefficients) for 1 MeV–100 GeV photons in elements $Z = 1$ to 100", *J. Phys. Chem. Ref. Data* **9**, 1023–1147.

Hubbell J.H., Wm.J. Veigele, E.A. Briggs, R.T. Brown, D.T. Cromer and R.J. Howerton (1975), "Atomic form factors, incoherent scattering functions, and photon scattering cross sections", *J. Phys. Chem. Ref. Data* **4**, 471–538. Erratum: ibid. **6** (1977) 615–616.

ICRU 37 (1984), *Stopping Powers for Electrons and Positrons* (ICRU, Bethesda, MD).

Inokuti M. (1971), "Inelastic collisions of fast charged particles with atoms and molecules–the Bethe theory revisited", *Rev. Mod. Phys.* **43**, 297–347.

Inokuti M., Y. Itikawa and J.E. Turner (1978), "Addenda: inelastic collisions of fast charged particles with atoms and molecules–the Bethe theory revisited", *Rev. Mod. Phys.* **50**, 23–35.

Inokuti M. and D.Y. Smith (1982), "Fermi density effect on the stopping power of metallic aluminum", *Phys. Rev. B* **25**, 61–66.

Jablonski A. (1987), "Effects of Auger electron elastic scattering in quantitative AES", *Surf. Science* **188**, 164–180.

Jackson J.D. (1975), *Classical Electrodynamics* (John Wiley and Sons, New York).

James F. (1980), "Monte Carlo theory and practice", *Rep. Prog. Phys.* **43**, 1145–1189.

James F. (1990), "A review of pseudorandom number generators", *Comput. Phys. Commun.* **60**, 329–344.

Jenkins T.M., W.R. Nelson and A. Rindi (1988), eds., *Monte Carlo Transport of Electrons and Photons* (Plenum, New York).

Kalos M.H. and P.A. Whitlock (1986), *Monte Carlo Methods*, vol. 1 (Wiley, New York).

Kane P.P., L. Kissel, R.H. Pratt and S.C. Roy (1986), "Elastic scattering of γ-rays and X-rays by atoms", *Phys. Rep.* **140**, 75–159.

Kawrakow I. and D.W.O. Rogers (2000), "The EGSnrc code system: Monte Carlo simulation of electron and photon transport", *Report PIRS-701* (National Research Council of Canada, Ottawa).

Kim L., R.H. Pratt, S.M. Seltzer and M.J. Berger (1986), "Ratio of positron to electron bremsstrahlung energy loss: an approximate scaling law", *Phys. Rev. A* **33**, 3002–3009.

Kirkpatrick P. and Wiedmann L. (1945), "Theoretical continuous X-ray energy and polarization", *Phys. Rev.* **67**, 321–339.

Kissel L., C.A. Quarles and R.H. Pratt (1983), "Shape functions for atomic-field bremsstrahlung from electrons of kinetic energy 1–500 keV on selected neutral atoms $1 \leq Z \leq 92$", *At. Data Nucl. Data Tables* **28**, 381–460.

Kittel C. (1976), *Introduction to Solid Physics* (John Wiley and Sons, New York).

Koch H.W. and J.W. Motz (1959), "Bremsstrahlung cross-section formulas and related data", *Rev. Mod. Phys.* **31**, 920–955.

Landau L. (1944), "On the energy loss of fast particles by ionisation", *J. Phys. U.S.S.R.* **8**, 201–207.

L'Ecuyer P. (1988), "Efficient and portable combined random number generators", *Commun. ACM* **31**, 742.

Lederer C.M. and V.S. Shirley (1978), eds., *Table of Isotopes*, 7th edition (Wiley, New York) appendix III.

Lewis H.W. (1950), "Multiple scattering in an infinite medium", *Phys. Rev.* **78**, 526–529.

Lewis H.W. (1952), "Range straggling of a nonrelativistic charged particle", *Phys. Rev.* **85**, 20–24.

Liljequist D. (1983), "A simple calculation of inelastic mean free path and stopping power for 50 eV–50 keV electrons in solids", *J. Phys. D: Appl. Phys.* **16**, 1567–1582.

Liljequist D. (1987), "Critical path length for the similarity of elastic multiple scattering processes", *J. Appl. Phys.* **62**, 333–341.

Lindhard J. and A. Winther (1964), "Stopping power of electron gas and equipartition rule", *Mat. Fys. Medd. Dan. Vid. Selsk.* **34**, 1–22.

Ljungberg M. and S.-E. Strand (1989), "A Monte Carlo program for the simulation of scintillation camera characteristics", *Comput. Meth. Programs Biomed.* **29**, 257–272.

Manson S.T. (1972), "Theoretical study of generalized oscillator strengths in atoms: comparison with experiment and other calculations", *Phys. Rev. A* **5**, 668–677.

Maron M.J. (1982), *Numerical Analysis: A Practical Approach* (Macmillan, New York).

Mayol R. and F. Salvat (1990), "Cross sections for K-shell ionisation by electron impact", *J. Phys. B: At. Mol. Opt. Phys.* **23**, 2117–2130.

Mohr P.J. and B.N. Taylor (2000), "CODATA recommended values of the fundamental physical constants: 1998", *Rev. Mod. Phys.* **72**, 351–495.

Molière G. (1947), "Theorie der Streuung schneller geladener Teilchen I. Einzelstreuung am abgeschirmten Coulomb-Feld", *Z. Naturforsch.* **2a**, 133–145.

Molière G. (1948), "Theorie der Streuung schneller geladener Teilchen II. Mehrfach- und Vielfachstreuung", *Z. Naturforsch.* **3a**, 78–97.

Møller C. (1932), "Zur Theorie des Durchgangs schneller Elektronen durch Materie", *Ann. Physik* **14**, 531–585.

Mott N.F. and H.S.W. Massey (1965), *The Theory of Atomic Collisions*, 3rd edition (Oxford Univ. Press, London).

Motz J.W., H.A. Olsen and H.W. Koch (1969), "Pair production by photons", *Rev. Mod. Phys.* **41**, 581–639.

Namito Y., S. Ban and H. Hirayama (1994), "Implementation of the Doppler broadening of a Compton-scattered photon into the EGS4 code", *Nucl. Instrum. Meth. A* **349**, 489–494.

Nelson W.R., H. Hirayama and D.W.O. Rogers (1985), "The EGS4 Code System", *Report SLAC-265* (Stanford Linear Accelerator Center, Stanford, CA).

Perkins S.T., D.E. Cullen, M.H. Chen, J.H. Hubbell, J. Rathkopf and J. Scofield (1991), "Tables and graphs of atomic subshell and relaxation data derived from the LLNL evaluated atomic data library (EADL), Z = 1–100", *Report UCRL-50400* vol. 30 (Lawrence Livermore National Laboratory, Livermore, CA).

Pratt R.H., A. Ron and H.K. Tseng (1973), "Atomic photoelectric effect above 10 keV", *Rev. Mod. Phys.* **45**, 273–325. Erratum: ibid. **45** (1973) 663–664.

Pratt R.H., H.K. Tseng, C.M. Lee and L. Kissel (1977), "Bremsstrahlung energy spectra from electrons of kinetic energy 1 keV $\leq T_1 \leq$ 2000 keV incident on neutral atoms $2 \leq Z \leq 92$", *At. Data Nucl. Data Tables* **20**, 175–209. Erratum: ibid. **26** (1981) 477–481.

Press W.H. and S.A. Teukolsky (1992), "Portable random number generators", *Computers in Physics* **6**, 522–524.

Reimer L. (1985), *Scanning Electron Microscopy* (Springer, Berlin).

Reimer L. and E.R. Krefting (1976), "The effect of scattering models on the results of Monte Carlo calculations", *National Bureau of Standards Special Publication 460* (US Government Printing Office, Washington DC) pp. 45–60.

Reimer L., U. Zepke, J. Moesch, St. Schulze-Hillert, M. Ross-Messemer, W. Probst and E. Weimer (1992), *EEL Spectroscopy* (Carl Zeiss, Oberkochen).

Ribberfors R. (1983), "X-ray incoherent scattering total cross sections and energy-absorption cross sections by means of simple calculation routines", *Phys. Rev. A* **27**, 3061–3070.

Ribberfors R. and K.-F. Berggren (1982), "Incoherent-x-ray-scattering functions and cross section $(d\sigma/d\Omega')_{incoh}$ by means of a pocket calculator", *Phys. Rev. A* **26**, 3325–3333.

Rubinstein R.Y. (1981), *Simulation and the Monte Carlo Method* (Wiley, New York).

Sakurai J.J. (1967), *Advanced Quantum Mechanics* (Addison and Wesley, New York).

Saloman E.B., J.H. Hubbell and J.H. Scofield (1988), "X-ray attenuation cross sections for energies 100 eV to 100 keV and elements $Z = 1$ to $Z = 92$", *At. Data Nucl. Data Tables* **38**, 1–197.

Salvat F. (1987), "Algorithms for random sampling from single-variate distributions", *Comput. Phys. Commun.* **46**, 427–436.

Salvat F. (1998), "Simulation of electron multiple elastic scattering", *Radiat. Phys. Chem.* **53**, 247–256.

Salvat F. (2000), Private communication.

Salvat F. and J.M. Fernández-Varea (1992), "Semiempirical cross sections for the simulation of the energy loss of electrons and positrons in matter", *Nucl. Instrum. Meth. B* **63**, 255–269.

Sauter F. (1931), "Über den atomaren Photoeffekt in der K-Schale nach der relativistischen Wellenmechanik Diracs", *Ann. Phys.* **11**, 454–488.

Schiff L.I. (1968), *Quantum Mechanics*, 3rd edition (McGraw-Hill Kogakusha Ltd., Tokyo).

Schultz P.J. and K.G. Lynn (1988), "Interaction of positron beams with surfaces, thin films, and interfaces", *Rev. Mod. Phys.* **60**, 701–770.

Seltzer S.M. and M.J. Berger (1985), "Bremsstrahlung spectra from electron interactions with screened atomic nuclei and orbital electrons", *Nucl. Instrum. Meth. B* **12**, 95–134.

Seltzer S.M. and M.J. Berger (1986), "Bremsstrahlung energy spectra from electrons with kinetic energy 1 keV–10 GeV incident on screened nuclei and orbital electrons of neutral atoms with $Z = 1$–100", *At. Data Nucl. Data Tables* **35**, 345–418.

Sempau J., E. Acosta, J. Baró, J.M. Fernández-Varea and F. Salvat (1997), "An algorithm for Monte Carlo simulation of coupled electron-photon transport", *Nucl. Instrum. Meth. B* **132**, 377–390.

Sevier K.D. (1972), *Low Energy Electron Spectrometry* (Wiley Interscience, New York).

Shiles E., T. Sasaki, M. Inokuti and D.Y. Smith (1980), "Self-consistency and sum-rule tests in the Kramers-Kronig analysis of optical data: applications to aluminum", *Phys. Rev. B* **22**, 1612–1628.

Snyder W.S., M.R. Ford, G.G. Warner and H.L. Fisher Jr. (1969), "Estimates of absorbed fractions for monoenergetic photon sources uniformly distributed in various organs of a heterogeneous phantom", MIRD Pamphlet No. 5, *J. Nucl. Med.* **10**, suppl. No. 3, 5–52.

Statham P.J. (1976), "The generation, absorption and anisotropy of thick-target bremsstrahlung and the implications for quantitative energy dispersive analysis", *X-Ray Spectrom.* **5**, 154–168.

Sternheimer R.M. (1952), "The density effect for the ionization loss in various materials", *Phys. Rev.* **88**, 851–859.

Sternheimer R.M., S.M. Seltzer and M.J. Berger (1982), "Density effect for the ionization loss of charged particles in various substances", *Phys. Rev. B* **26**, 6067–6076. Erratum: ibid. **27** (1983) 6971.

Titus F. (1970), "Measurements of the energy response functions of scintillators for monoenergetic electrons", *Nucl. Instrum. Meth.* **89**, 93–100.

Tofterup A.L. (1986), "Theory of elastic and inelastic scattering of electrons emitted from solids: energy spectra and depth profiling in XPS/AES", *Surf. Science* **167**, 70–100.

Tsai Y.S. (1974), "Pair production and bremsstrahlung of charged leptons", *Rev. Mod. Phys.* **46**, 815–851.

Walker A.J. (1977), "An efficient method for generating discrete random variables with general distributions", *ACM Trans. Math. Software* **3**, 253–256.

Walker D.W. (1968), "Spin polarization in electron scattering from molecules", *Phys. Rev. Lett.* **20**, 827–828.

Walker D.W. (1971), "Relativistic effects in low energy electron scattering from atoms", *Adv. Phys.* **20**, 257–323.

Wentzel G. (1927), "Zwei Bemerkungen über die Zerstreuung korpuskularer Strahlen als Beugungserscheinung", *Z. Phys.* **40**, 590–593.

Yates A.C. (1968), "Calculations of electron spin polarization for electron-molecule collisions", *Phys. Rev. Lett.* **20**, 829–831.

Zerby C.D. (1963), "A Monte Carlo calculation of the response of gamma-ray scintillation counters", in *Methods in Computational Physics*, vol. 1, eds. B. Alder, S. Fernbach and M. Rotenberg (Academic Press, New York) pp. 89–134.

Zheng-Ming L. and A. Brahme (1993), "An overview of the transport theory of charged particles", *Radiat. Phys. Chem.* **41**, 673–703.

PENELOPE – A Code System for Monte Carlo Simulation of Electron and Photon Transport

Radiation is used in many applications of modern technology. Its proper handling requires competent knowledge of the basic physical laws governing its interaction with matter. To ensure its safe use, appropriate tools for predicting radiation fields and doses, as well as pertinent regulations, are required.

One area of radiation physics that has received much attention concerns electron-photon transport in matter. PENELOPE is a modern, general-purpose Monte Carlo tool for simulating the transport of electrons and photons, which is applicable for arbitrary materials and in a wide energy range. PENELOPE provides quantitative guidance for many practical situations and techniques, including electron and x-ray spectroscopies, electron microscopy and microanalysis, biophysics, dosimetry, medical diagnostics and radiotherapy, and radiation damage and shielding.

These proceedings contain the teaching notes of a recent workshop/training course on PENELOPE, with a detailed description of the physics, numerical algorithms and structure of the code system.

ISBN 92-64-18475-9

OECD Nuclear Energy Agency
Le Seine Saint-Germain – 12, boulevard des Îles
F-92130 Issy-les-Moulineaux, France
Tel.: +33 (0)1 45 24 10 15 – Fax: +33 (0)1 45 24 11 10
E-mail: nea@nea.fr – Internet: www.nea.fr

 670

A modified version of the DPM Monte Carlo code for radiotherapy treatment planning inside of uniform magnetic fields

James F. Dempsey, Ph.D.[1,2] and Ricardo A. Yanez, Ph.D.[2]

[1]Department of Radiation Oncology, University of Florida, Gainesville, Florida 32610, U.S.A.
[2]ViewRay Incorporated, Gainesville, Florida 32601, U.S.A.

Corresponding author & for reprints:
James F. Dempsey, Ph.D.
Department of Radiation Oncology
University of Florida College of Medicine
P.O. Box 100385
Gainesville, FL 32610-0385
Phone: (352) 265-8217
FAX: (352) 265-8417
E-Mail: dempsey@ufl.edu

Acknowledgments: This research was supported in part by NIH grant No. R01 CA100636 awarded by the National Cancer Institute and by the National Science Foundation under Grant No. DMI-0457394.

Conflict of Interest Notification:

James F. Dempsey owns stock in and is the Chief Science Officer and cofounder of ViewRay, Inc., Suite 201-D, 101 S.E. 2nd Place, Gainesville, FL 32601, phone: 352-374-4005, fax: 352-380-0600, http://www.viewray.com. ViewRay, Inc. is a North Florida based company dedicated to the advancement of the field of Radiation Oncology through the development of state-of-the-art image guided intensity modulated radiation therapy technology. As such, he may benefit financially as a result of the outcomes of work or research reported in this manuscript.

Abstract:

We present studies in support of the development of a magnetic resonance imaging (MRI) guided intensity modulated radiation therapy (IMRT) device for the treatment of cancer patients. Fast and accurate computation of the absorbed ionizing radiation dose delivered in the presence of the MRI magnetic field are required for clinical implementation. The fast Monte Carlo simulation code DPM, optimized for radiotherapy treatment planning, was modified to simulate electron transport in uniform, static magnetic fields. Simulations of dose deposition in inhomogeneous phantoms in which a layer of air is sandwiched in water shows that a lower MRI field strength is to preferred in order to avoid dose build-up, due to returning electrons, in tissue surfaces. At 1.5 T, large dose build up is observed in both lung tissue and air. Even a small 2 mm air cavity can produce a large dose build up at 1.5 T. For a magnetic field of 0.3 T, dose build up in lung is insignificant while cavities of up to approximately ~1 cm in thickness are safe from large dose build-ups. Larger cavities, of 2 cm or greater, do display some build up that will need to be accurately modeled and accounted for in treatment planning, but are of a much lower magnitude than at high field strength.

Key Words: Monte Carlo; MRI; IGRT

1. INTRODUCTION

An active program to commercially develop an integrated Co-60 intensity modulated radiation therapy (IMRT) unit with a superconducting magnetic resonance imaging (MRI) unit, the development of the Renaissance™ System, has prompted for the advance of fast, reliable and accurate dose treatment planning calculations in targets submerged in uniform, static, magnetic fields. The electron transport Monte Carlo method is becoming an important tool for clinical radiotherapy dose planning because of its superior accuracy compared to deterministic models. Its routine use in clinical applications has thus far been hindered by the comparably long computational times. However, the inherent independent nature of radiation transport events, which makes the method strictly parallel, offers an appealing approach towards achieving computational times consistent with clinical use. Recent efficiency improvements of transporting algorithms, like the one implemented in the Dose

Planning Method (DPM) code [1] by combining step-size independent multiple scattering theory [2], the "random hinge" technique and the Woodcock photon tracking scheme, offers additional computational time improvements.

The purpose of this work is to advance in the feasibility demonstration of integrating a MRI unit with a Co-60 IMRT unit. By measuring and assessing each individual cancer patient's actual delivered daily dose, patient-motion related dose-delivery errors, which include daily patient set-up errors, long term physiological changes, and internal organ motions, are minimized. The MRI scanner is designed to be a low field unit (~ 0.3 T) to allow for imaging with spatial integrity by limiting magnetic susceptibility artifacts due to the patient and to prevent significant perturbations of the dose distribution.

The effect of uniform, static magnetic fields on simulated absorbed doses has been studied with a modified version of DPM, and the results have been compared to PENELOPE [3, 4]. We have chosen DPM because it is optimized for clinical treatment planning, and because of its computational speed advantage over other standard Monte Carlo radiation transport codes. Routine Monte Carlo dose planning calculations will be required to achieve optimal dose delivery, and a high performance computer for parallel processing is planned to be an integrated part of Renaissance to achieve this goal. DPM has been benchmarked against EGS4 [5], a code that has been thoroughly tested in the energy range relevant for dosimetric applications.

2. MATERIALS AND METHODS

2.1 RADIATION TRANSPORT WITH MAGNETIC FIELDS

The motion of a particle of charge q and velocity $\vec{v}$ in the presence of a magnetic field $\vec{B}$, set up by external sources, is described by the well-known Lorentz force equation (in Gaussian units),

$$\frac{d\vec{p}}{dt} = \frac{q}{c}\vec{v} \times \vec{B} \qquad (1)$$

where $\vec{p}$ is the particle momentum and c is the magnitude of the velocity of light in vacuum. If the energy is constant, as is the case when the charged particle moves in vacuum, the magnitude of the velocity is constant, and Eq. (1) can be written,

$$\frac{d\vec{v}_{\parallel}}{dt} = 0, \qquad \frac{d\vec{v}_{\perp}}{dt} = \vec{v}_{\perp} \times \vec{\varpi}_B \tag{2}$$

where,

$$\vec{\varpi}_B = \frac{q\vec{B}}{\gamma mc} \tag{3}$$

is the precession frequency, and $\vec{v}_{\parallel} = (\vec{v} \cdot \hat{\omega})\hat{\omega}$ and $\vec{v}\perp = \vec{v} - (\vec{v} \cdot \hat{\omega})\hat{\omega}$ are the parallel and perpendicular components of $\vec{v}$ relative to $\vec{\varpi}_B$, respectively. Eq. (2) describes a uniform translation parallel to $\vec{B}$ and a circular motion perpendicular to $\vec{B}$ with angular velocity $\vec{\varpi}_B$ and gyration radius,

$$R_B = \frac{v_{\perp}}{\omega_B} \tag{4}$$

In each interaction event simulated with DPM, where electrons ($q = -e$) or positrons ($q = +e$) are involved, the equations of motion, Eq. (2), are solved by integration, assuming the energy of the particle is not changed throughout the step. This simplification ignores the fact that ionizing charged particles in fact loose energy continuously in the medium. However, within the simulated step, the energy loss is small and it is assumed that energy is for all practical purposes fixed, and Eq. (2) is valid as long as the integration is performed within that step.

The more exact equations of motion in a magnetic field derived by D. Jette [6], which include an explicit term describing the energy loss in the medium,

$$\frac{d\vec{p}}{dt} = \frac{q}{c}\vec{v} \times \vec{B} - S(E(s))\hat{v} \tag{5}$$

where S is a continuous function describing the continuous slowing down approximation (CSDA) energy loss along the trajectory, s. In the Monte Carlo simulation, multiple scattering theory is used to simulate the global effect of a large number of events in a track segment of a given length. The equations of motion in Ref. [6]

reduce to Eq. (2) when solved within such a segment. In our approach, we have implemented the equations of motion Eq. (5). By assuming the energy loss has the simple CSDA form:

$$S(E) = -\frac{dE}{ds} \tag{6}$$

An approximate solution to Eq. (5) can be derived for a homogeneous medium if it is assumed the charged particle energy, above a threshold, decreases linearly with depth,

$$E(s) = S_T(r_o - s) \tag{7}$$

where S_T is a constant and r_o is the particle range. The magnetic field is along the y-direction and the direction of the velocity is written for convenience,

$$\hat{v} = \cos(\xi)\sin(\eta\hat{x}) + \sin(\xi\hat{y}) + \cos(\xi)\cos(\eta\hat{z}) \tag{8}$$

This choice makes the angle ξ become a constant of the motion, $\xi = \xi_o$. Under the assumption given by Eq. (7), the angle η is then given by,

$$\eta(s) = \eta_o + \zeta\left(\frac{s}{r_o}\right) \tag{9}$$

where,

$$\zeta(\tau) = \frac{qB}{S_T}\ln\left[\frac{\sqrt{\alpha(\alpha+2)}+\alpha+1}{\sqrt{\alpha(1-\tau)\left[\alpha(1-\tau)+2\right]}+\alpha(1-\tau)+1}\right] \tag{10}$$

$$\alpha = \frac{S_T r_o}{m_e c^2} \tag{11}$$

and η_o is the initial angle. If (x_o, y_o, z_o) is the initial position, then the trajectory is,

$$x(s) = x_o + r_o\cos(\xi_o)\left[\sin(\eta_o)P\left(\frac{s}{r_o}\right) + \cos(\eta_o)Q\left(\frac{s}{r_o}\right)\right] \tag{12}$$

$$y(s) = y_o + \cos(\xi_o)s \tag{13}$$

$$z(s) = z_o + r_o\cos(\xi_o)\left[\cos(\eta_o)P\left(\frac{s}{r_o}\right) - \sin(\eta_o)Q\left(\frac{s}{r_o}\right)\right] \tag{14}$$

where,

$$P(\tau) = \int_0^\tau \sin(\zeta(\tau'))d\tau' \quad (15)$$

$$Q(\tau) = \int_0^\tau \cos(\zeta(\tau'))d\tau' \quad (16)$$

At each step, $S_T = E_o / r_o$ is evaluated, where E_o is the initial kinetic energy of the particle, and Eqs. (12) are solved numerically. The outcome is found to be nearly identical to Eq. (2) when solved within a step that preserves the accuracy of a simulation. Even when combining several contiguous steps together, the solution of the equations of motion in Ref. [6] involve the numerical evaluation of integrals, Eq. (13) and Eq. (14).

2.2 SIMULATIONS WITH CO-60 PHOTONS

In radiation transport, only the transport of charged particles is directly affected by the magnetic field. In a photon beam, electrons and positrons emerge as a result of photon interactions in the material, a fraction, or the total, of the incoming energy being transferred, and suffer perceptible defections due to the magnetic field only when the mean free path in the material is of the order of, or larger than, the gyration radius, Eq. (4). This condition is met in a variety of situations, depending on the energy of the primary photon, which determines the velocity spectrum of secondary electrons, the magnetic field strength, and the density and material composition. In the case of Co-60 photons, high magnetic fields cause sizable effects in the dose deposition in tissue in the build-up region.

Material interfaces, where an abrupt change in density and mean free path may occur, have been identified as potentially prone to dose distribution perturbations. In the following set of simulations we model a primary photon beam of Co-60, of infinitesimal cross sectional area, impinging normally upon an inhomogeneous phantom of $10 \times 10 \times 10$ cm^3 ($200 \times 200 \times 200$ voxels), consisting of water with a slab of 2 cm thick lung tissue (ρ=0.3 g/cm3) in the center. A coordinate system in which the z-axis is aligned with the incident beam direction is adopted. In this coordinate system, a uniform, static magnetic field B is assumed to be aligned along the y

direction. Calculations were performed with no magnetic field, and with $B=0.3$ T and $B=1.5$ T. The maximum stepsize was $smax=0.1$ cm, and cut-off energies for inelastic collisions W_{cc} = 200 keV and W_{br} =50 keV for bremsstrahlung emission, respectively, were the same for both materials. Fig. 1 shows the integrated depth dose per primary particle, which for an infinitesimal beam is almost equivalent to the central axis depth dose, for no magnetic field (blackline), 0.3T(redline) and1.5T(blueline), simulated with DPM.

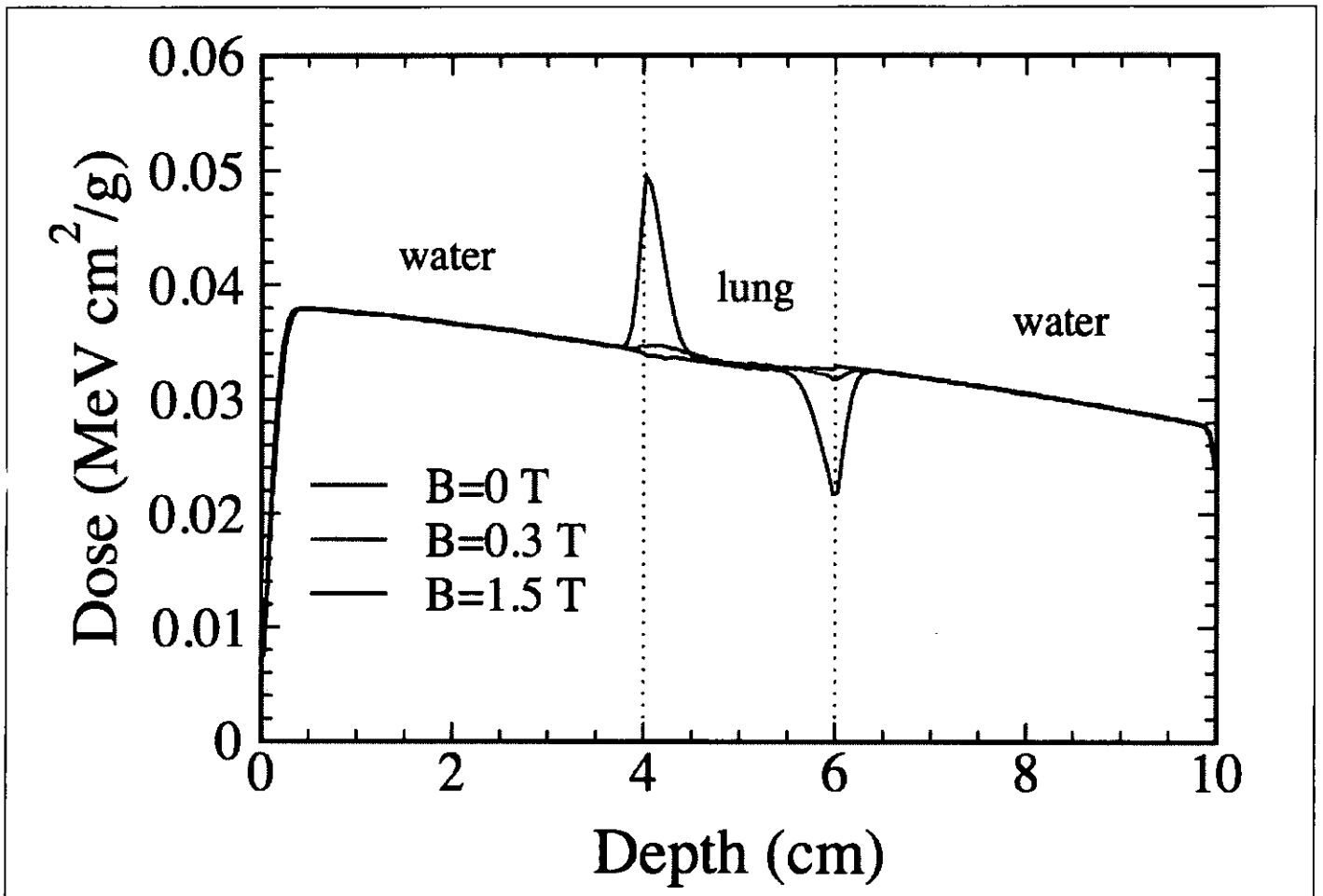


FIG. 1: Integrated depth dose in a water phantom with a 2cm slab of lung density tissue in the center with no magnetic field (black), B=0.3T (red), and B=1.5T (blue), respectively. Note that significant errors only exist at 1.5 T.

At B=0.3T the mean free path of secondary electrons is small compared to the gyration radius, and electrons follow soft curvilinear trajectories while loosing energy in the material. At a field strength of B=1.5T, a dose build-up is formed around the first boundary. The shorter gyration radius produced by the higher field force

some electrons to spin back, depositing energy near the intersection of the two materials. A corresponding dose depletion, due to the loss of equilibrium, is observed around the second boundary. As demonstrated in Ref. [7], an opposing beam tends to cancel out dose deformations due to magnetic fields; a dose enhancement in one boundary produced by a beam is a dose depletion produced by the opposing beam, and visa versa. The superposition of the two cancel the effect, as shown in Fig. 2a, where two opposing, but otherwise identical, beams have been used in the $B=1.5\ T$ case. In a symmetric phantom with a centered cavity, the cancellation is expected to be almost complete. In an asymmetric phantom, the cancellation using identical opposing beams is far from perfect, as shown in Fig. 2b. The dashed and dotted lines show the depth doses for the beams entering from the front and backsides of the phantom, respectively, while the solid line is the sum. By integrating the peak and valley around the surfacea $t_z=4\ cm$, and using the ratio to modulate the intensities of the front and back beams, minimization of the dose enhancement at the material interface can be achieved. Fig.2c shows a simulation in which the intensity of the opposing beam is three times more intense than the front-facing beam, corresponding to an intensity ratio, of ~0.25.

 3679

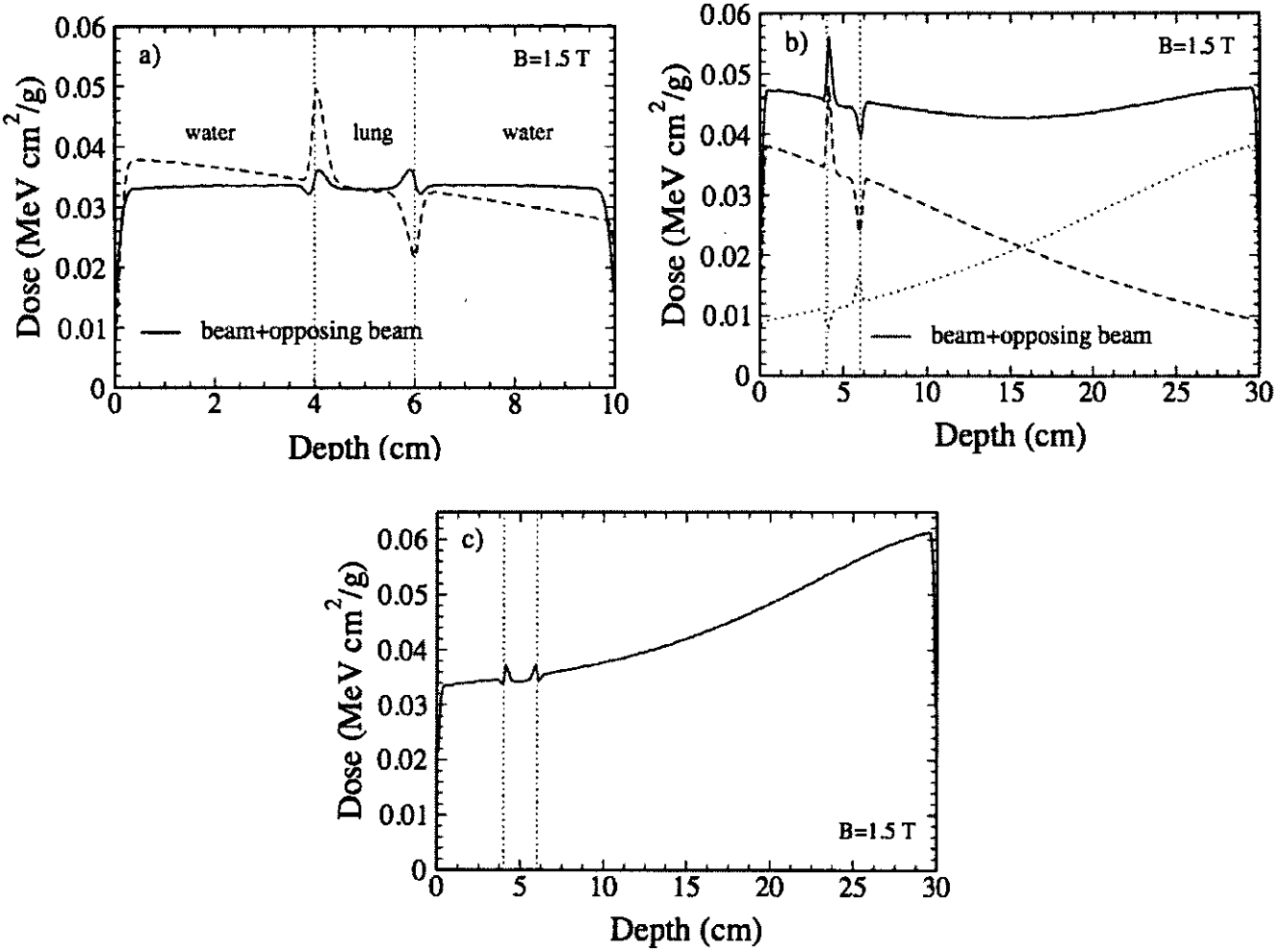


FIG. 2: a) Integrated depth dose in a water phantom with a 2 cm lung tissue slab in the center for B=1.5T, normal beam (dashed line) and using opposing beams (solidline). b) in an asymmetric 10×10×30 cm^3 water phantom with a 2 cm lung tissue slab centered around z =5cm. c) modulating the beam intensities according to peak and valley ratios.

2.3 SIMULATIONS WITH LUNG & AIR CAVITIES.

The original DPM is not meant to deal, in an efficient way, with combinations of materials with significantly different densities, as in the case of inhomogeneous phantoms made of water (ρ = 1.0 g/cm^3) and air (ρ = ρ = 0.0012 g/cm^3). The main limitation of the original code is the use of a common code-wide cut-off energy for all materials. For electrons, the energy equivalent of a range of 1 mm in water is ~ 350 keV, whereas

this equivalent energy is ~ 10 keV in air. Hence, if the side of a voxel is 1 mm, the transport should be simulated with cut-off energies for inelastic collisions of the order of ~ 350 keV in water, and ~ 10 keV in air. In this situation, a common cut-off energy of ~ 10 keV for electrons in water and air seem unjustified, slowing the simulations down considerably. Moreover, as the energy cut-offs are lowered, the maximum step-size should be decreased accordingly. Therefore, several modifications to DPM where made in order to accommodate a per material Wcc , Wbr and smax .

In the following set of simulations we model a primary photon beam of Co-60, of 4 × 4 cm^2 cross-sectional area, impinging normally upon a inhomogeneous phantom of 10 × 10 × 10 cm^3 (200 × 200 × 200 voxels). In Fig. 3 a two-dimensional map of the depth dose, integrated between a slab of +/- 1 mm in the y direction from the center axis of the phantom, is shown for an inhomogeneous phantom consisting of water with a 2 cm gap of air in the center.

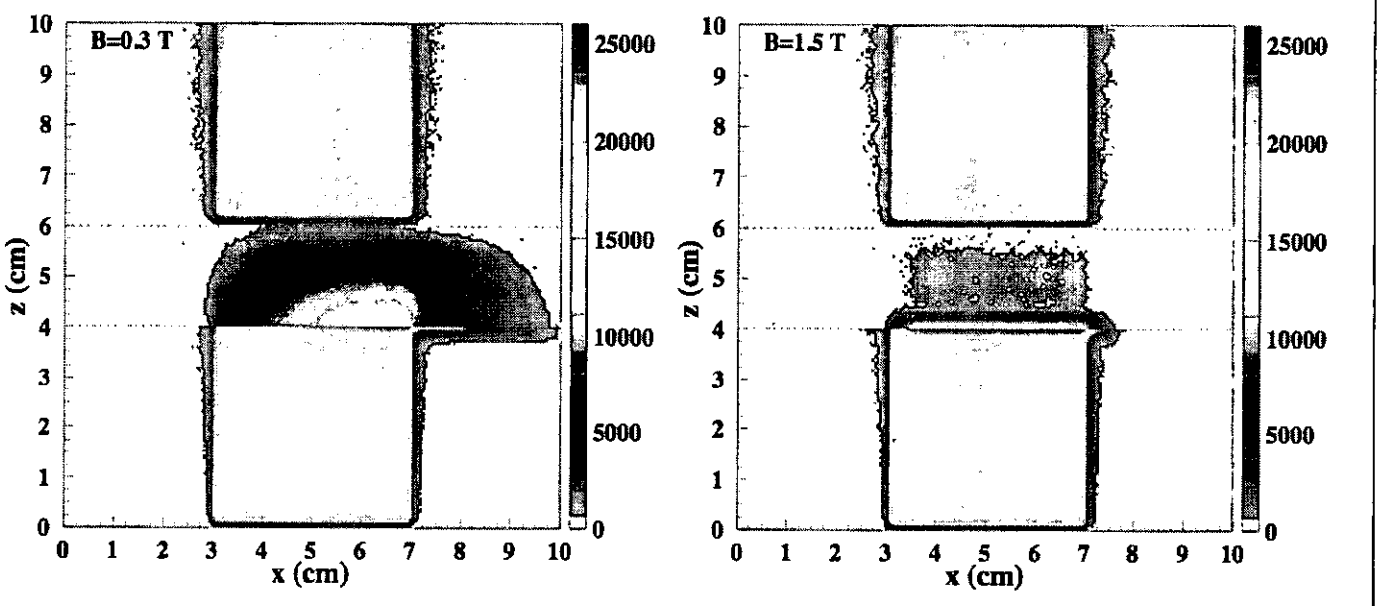


FIG. 3: 2-D dose depth map in the x-z plane for a water phantom with a large 2 cm gap of air in the center for B = 0.3 T and B = 1.5 T, respectively.

This set of simulations were made with the improved version of DPM in which the simulation parameters for water were smax =0.5cm, and cut-off energies for inelastic collisions Wcc =200 keV and Wbr = 50 keV for

bremsstrahlung emission, respectively. For air it was found that values of smax=0.05 cm and Wcc = 20keV preserved the accuracy of the simulation. A dose build-up is formed on the boundary surface due to returning electrons, even for the lower field, though with significantly diminished magnitude. The reason is that the gap is larger than the largest gyration radius. The long gyration radius of electrons at B=0.3T suggest that a narrower gap, smaller than the largest gyration radius, will cause some electrons to hit and be absorbed in the second surface, instead of returning back to the first. This is shown in Fig. 4. where the air gap is 2mm.

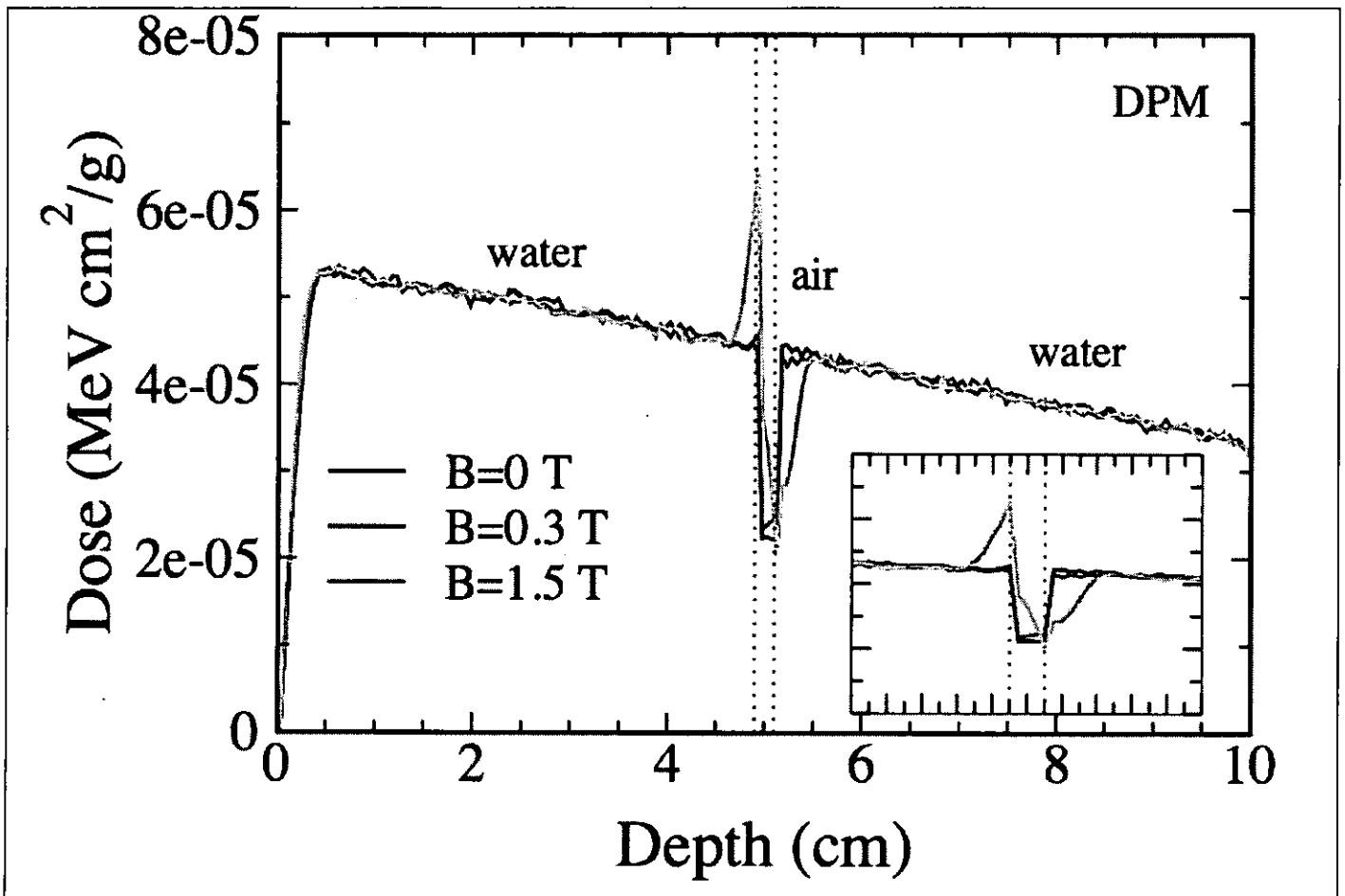


FIG. 4: Central axis depth dose in a water phantom with a small 2 mm air gap in the center with no magnetic field, and B=0.3T and B=1.5T, respectively. Inset shows the depth dose at a depth between 4 and 6 cm.

At B=0.3T no build-up is observed. In contrast, a significant build-up is developed at B=1.5T even for this very small gap, because the typical gyration radius is still shorter than the gap. As the dimension of the cavity

 3632

becomes larger than the largest gyration radius, all electrons entering air return back to the phantom. For B=1.5 T this happens for cavities larger than 3-4 mm, and for B=0.3 T for cavities larger than ~1.3 cm.

2.4 COMPARISON AND VALIDATION WITH PENELOPE.

In Fig. 5 we show a set of repeated simulations obtained with our modified version of DPM and PENELOPE using similar simulation parameters as the DPM simulation, and C1=C2=0.02. In Fig. 7 the DPM and PENELOPE central axis depth dose, integrated between ±1 mm in the x- and y-direction, is shown for no field, and B = 1.5 T, respectively. Both sets of simulations are for any practical purpose identical, except for computation speed. On a 2.0 GHz 64-bit Intel Xeon processor, PENELOPE simulates the B = 0.3 T case at a rate of ~ 5k histories/s, whereas DPM does the same simulation at a rate of ~ 200k histories/s, ~40 times faster than PENELOPE.

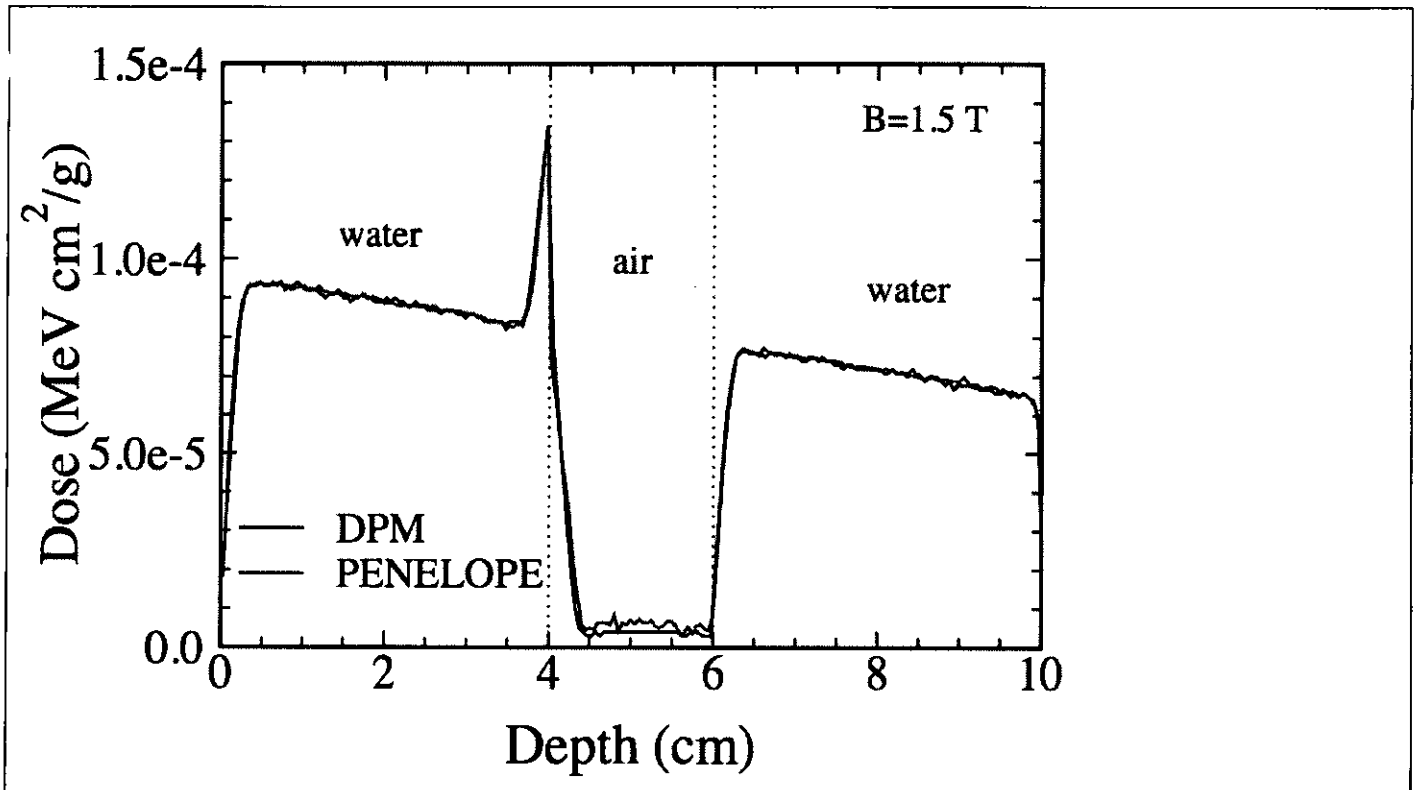


FIG. 7: Central axis depth dose in a water phantom with a 2 cm air gap in the center with B = 1.5 T.

3. CONCLUSIONS

We have successfully modified DPM to simulated the effect on the doses in an environment where uniform, static magnetic fields are present, yet keeping the computational time advantage of DPM over other standard Monte Carlo codes used in radiotherapy dose calculations. The effect of a B = 0.3 T magnetic field in the dose deposition by a Co-60 source is only appreciable in large air cavities, where a dose build-up due to returning electrons is formed near material boundaries. Cavities with dimensions smaller than the gyration radius are not severely affected by dose build-ups. For B = 0.3 T cavities of the order of ~ 1 cm are safe from dose build-ups.

REFERENCES

[1] A.F. Biela jew J. Sempau, S.J. Wilderman. Dpm, a fast, accurate monte carlo code optimized for photon and electron radiotherapy treatment planning dose calculations. Phys. Med. Biol., 45:2263, 2000.

[2] A.F. Biela jew I. Kawrakow. On the representation of electron multiple elastic-scattering distributions for monte carlo calculations. Nucl. Inst. Meth. Phys. Res. B, 134:325, 1998.

[3] J. M. Fernndez-Varea J. Bar and F. Salvat. Penelope: An algorithm for monte carlo simulation of the penetration and energy loss of electrons and positrons in matter. Nucl. Inst. Meth. Phys. Res. B, 100:31, 1995.

[4] E. Acosta J. Sempau F. Salvat, J. M. Fernandez-Varea. Penelope - a code system for monte carlo simulation of electron and photon transport. Technical report, Paris: NEA-OECD, 2001.

[5] H. Hirayama W.R. Nelson and D.W.O. Rogers. The egs4 code system. Technical Report Report SLAC-265, Stanford Linear Accelerator Center, 1985.

[6] D. Jette. Magnetic fields with photon beams: Dose calculation using electron multiple-scattering theory. Med. Phys., 27(8):1705, 2000.

[7] B.W. Raaymakers A.J.E. Raaijmakers and J.J.W. Lagendijk. Integrating a mri scanner with a 6 mv radiotherapy accelerator: dose increase at tissue-air interfaces in a lateral magnetic field due to returning electrons. Phys. Med. Biol., 50:1363, 2005.

[8] I.J. Chetty N. Tyagi, A. Bose. Implementation of the dpm monte carlo code on a parallel architecture for treatment planning applications. Med. Phys., 31(9):2721, 2004.

[9] Renaissance(TM) System 1000, ViewRay, Incorporated. http://www.viewray.com/

Magnetic fields with photon beams: Dose calculation using electron multiple-scattering theory

David Jette[a)]
The Lawrence H. Lanzl Institute of Medical Physics, P.O. Box 30760, Seattle, Washington 98103-0760 and Department of Medical Physics, Rush-Presbyterian-St. Luke's Medical Center, Chicago, Illinois

(Received 3 November 1999; accepted for publication 10 May 2000)

Strong transverse magnetic fields can produce large dose enhancements and reductions in localized regions of a patient under irradiation by a photon beam. We have developed a new equation of motion for the transport of charged particles in an arbitrary magnetic field, incorporating both energy loss and multiple scattering. Key to modeling the latter process is a new concept, that of "typical scattered particles." The formulas which we have arrived at are particularly applicable to the transport of, and deposition of energy by, Compton electrons and pair-production electrons and positrons generated within a medium by a photon beam, and we have shown qualitatively how large dose enhancements and reductions can occur. A companion article examines this dose modification effect through systematic Monte Carlo simulations. © *2000 American Association of Physicists in Medicine.* [S0094-2405(00)00408-9]

Key words: photon dose calculation, magnetic fields, energy loss, multiple scattering

I. INTRODUCTION

The effect of transverse magnetic fields on the dose distributions for electron beams has been studied by a number of investigators since 1975. Shih[1] carried out Monte Carlo calculations including the effects of multiple scattering and energy straggling, and found that magnetic fields of 6 T (tesla) prevented the spreading out of the electrons, thereby obtaining a dose peak at the end of their range. Then Whitmire *et al.*[2–4] obtained experimental dose distributions for polystyrene and cork phantoms for magnetic fields of 0.9–1.8 T, and found a significant reduction in surface dose, a 20% increase in absolute dose in the maximum-dose region, and a sharper falloff towards the end of the range. Subsequently Paliwal *et al.*[5–7] showed that magnetic fields of 0.5–1 T reduced dose perturbations caused by tissue inhomogeneities. Finally, Nath[8] confirmed a large dose enhancement in the maximum-dose region and a steeper falloff towards the end of the range.

The next subject of investigation was the effect of longitudinal magnetic fields on electron-beam dose distributions. Using a field strength of 3 T, Nath[8] found a lower entrance dose, sharper beam edges, and a large increase in peak dose, and Weinhous *et al.*[9] subsequently carried out Monte Carlo calculations with a magnetic field strength of 1–4 T to systematically confirm these effects, pointing out the potential for sparing intervening tissue provided by the great increase in maximum dose relative to surface dose. Recently Bielajew[10] has carried out EGS4[11] Monte Carlo investigations for field strengths of 3 and 20 T, finding sharply reduced penumbral width and greatly reduced dose perturbations distal to tissue inhomogeneities for the stronger, but not the weaker, of these two fields. He also showed that for a broad parallel electron beam, the central-axis dose distribution is unaffected by the magnetic field, but that for a divergent (point) source the enhancement of the maximum dose does occur.

Very recently, Nardi and Barnea[12] have carried out Monte Carlo simulations of the dose deposited by 15-MeV electron beams in the presence of uniform transverse magnetic fields of strength 1–9 T which are applied suddenly after the electrons have penetrated 4 cm into the tissue equivalent material. Their study is similar to ours in that we are concerned with the effect of a transverse magnetic field on the secondary charged particles (electrons and positrons) in a photon beam. They found very large dose enhancements for magnetic fields of 3 T, and dose enhancements which were less pronounced for 2 T and of little benefit for 1 T. In the present study we shall find comparable results at the theoretical level.

Discussion of the effects of magnetic fields on photon-beam dose distributions is largely absent from the literature, although Bielajew[10] did find sharply reduced penumbral width for very strong longitudinal magnetic fields. However, Reiffel[13] has recently proposed the use of topical magnets to generate localized gradients in strong transverse magnetic fields to alter dose distributions from clinical photon beams. In a collaboration between his Exelar Corporation and the Medical Physics Department of Rush-Presbyterian-St. Luke's Medical Center in Chicago, possible applications of this idea are being investigated using Monte Carlo simulations as well as dosimetry studies using phantoms and specially designed small high-performance superconducting magnets.[14]

As part of the foregoing project, Reiffel, Li, and Chu have carried out EGS4 Monte Carlo calculations for transverse magnetic fields with strong longitudinal gradients, demonstrating the possibility of using strong transverse magnetic fields to obtain very large dose enhancements in localized regions of a patient irradiated by a photon beam. This phe-

 0094-2405/2000/27(8)/1705/12/$17.00

nomenon results from the increased curling of the secondary electrons and positrons about magnetic field lines for increased magnetic field strength; thus there is a major dose buildup right after the photons have passed through a region of sharply increasing magnetic field strength.

The purpose of this article is to develop the theory needed to systematically investigate the effects of strong transverse magnetic fields on dose distributions from photon beams. Section II incorporates energy loss into the equation of motion for a charged particle under the influence of a magnetic field, and Sec. III applies this theory to the secondary electrons and positrons in a photon beam, for a uniform transverse magnetic field. The effect of multiple scattering is incorporated into the equation of motion in Sec. IV, based on the new concept of "typical scattered particles," and general formulas are developed for a uniform transverse magnetic field. Finally, in conclusion, Sec. V emphasizes that the foregoing theoretical work only lays the basis for systematic research to develop a new modality which can be expected to provide significant improvement in radiation treatment for certain types of cancer.

II. ENERGY LOSS

In the Gaussian system of units, the equation of motion for a particle of charge e and of velocity $\vec{v}$ in a magnetic field $\vec{B}$ is[15]

$$\frac{d\vec{p}}{dt}=\frac{e}{c}\vec{v}\times\vec{B}, \tag{1}$$

where c is the speed of light. The momentum $\vec{p}$ can be written as

$$\vec{p}=\gamma m\vec{v}, \tag{2}$$

where m is the mass of the particle and

$$\gamma\equiv\frac{1}{\sqrt{1-\frac{v^2}{c^2}}}. \tag{3}$$

Let s denote distance along the trajectory of the particle, so that

$$v=\frac{ds}{dt}. \tag{4}$$

Then, using the identity

$$\frac{\vec{v}\cdot\vec{v}}{c^2}\gamma^2=\gamma^2-1, \tag{5}$$

we readily deduce

$$\frac{d\vec{p}}{dt}=mc^2\frac{d\gamma}{ds}\hat{v}+m\gamma v^2\frac{d\hat{v}}{ds}, \tag{6}$$

where $\hat{v}$ is a unit vector in the direction of the particle's motion: $\vec{v}=v\hat{v}$. Therefore Eq. (1) is equivalent to the following pair of equations:

$$mc^2\frac{d\gamma}{ds}\hat{v}=\vec{0} \tag{7}$$

and

$$m\gamma v^2\frac{d\hat{v}}{ds}=\frac{e}{c}\vec{v}\times\vec{B}. \tag{8}$$

Now let E represent the kinetic energy of the particle, so that

$$\gamma mc^2=E+mc^2. \tag{9}$$

Clearly Eq. (7) is a statement that the particle is losing no energy while in the magnetic field. But suppose that the particle does lose energy because it is traveling within a material and is continually colliding with atoms of that material, according to the continuous slowing-down approximation,

$$\frac{dE}{ds}=-S_T(E), \tag{10}$$

where S_T is the total (linear) stopping power. Equation (10) determines E as a function of s, for if the initial energy of the particle is E_0, we have

$$\int_{E(s)}^{E_0}\frac{dE}{S_T(E)}=s. \tag{11}$$

Thus instead of Eq. (7), we want

$$mc^2\frac{d\gamma}{ds}\hat{v}=-S_T(E(s))\hat{v}, \tag{12}$$

so that we should modify Eq. (1) to read

$$\frac{d\vec{p}}{dt}=-S_T(E(s))\frac{\vec{v}}{v}+\frac{e}{c}\vec{v}\times\vec{B}. \tag{13}$$

We can consider Eq. (13) to be the equation of motion of a particle traveling in a material under the influence of a magnetic field, losing energy to the material according to the continuous slowing-down approximation. We have not yet included multiple scattering, which results in additional lateral deflection, in our description of the particle's motion, but in Sec. IV of this article this important process will be incorporated into Eq. (13).

As we have seen, the longitudinal component (parallel to $\vec{v}$) of Eq. (13) is equivalent to Eq. (12). With the use of Eqs. (5) and (9), the transverse component (perpendicular to $\vec{v}$), Eq. (8), becomes

$$\sqrt{E(s)[E(s)+2mc^2]}\frac{d\hat{v}}{ds}=e\hat{v}\times\vec{B}, \tag{14}$$

with $E(s)$ defined by Eq. (11). Equation (14) is the general equation to be solved, for whatever magnetic field $\vec{B}(x,y,z)$ is specified to exist; this leads to a unique determination of $\hat{v}$ as a function of s, and thereby to the particle trajectory itself (given the initial location and direction of the particle). In order to get a handle on what sort of trajectories may result, let us in the remainder of this section consider a uniform magnetic field in the y-direction,

$$\vec{B} = B\hat{y}, \tag{15}$$

with B a constant.

We use the angles $\xi(s)$ and $\eta(s)$ to specify $\hat{v}(s)$:

$$\hat{v} = \cos\xi \sin\eta \ \hat{x} + \sin\xi \hat{y} + \cos\xi \cos\eta \ \hat{z}. \tag{16}$$

(This is a nonstandard spherical-coordinate representation which is mathematically useful here. Geometrically, ξ is the angle between $\hat{v}$ and its projection $\hat{v}_{\text{proj}}$ onto the x-z plane, and η is the angle between $\hat{v}_{\text{proj}}$ and the z axis.) The y-component of Eq. (14) then shows that ξ is a constant,

$$\xi = \xi_0, \tag{17}$$

where ξ_0 is the initial value of ξ. Except for the trivial case $\xi_0 = \pi/2$, the x- and z-components of Eq. (14) then give

$$\sqrt{E(s)[E(s)+2mc^2]}\frac{d\eta}{ds} = -eB, \tag{18}$$

whence

$$\eta(s) = \eta_0 - eB\int_0^s \frac{ds'}{\sqrt{E(s')[E(s')+2mc^2]}}, \tag{19}$$

where η_0 is the initial value of η. Thus, if the initial position of the particle is (x_0, y_0, z_0), its trajectory is

$$x(s) = x_0 + \cos\xi_0 \int_0^x \sin[\eta(s')]ds',$$

$$y(s) = y_0 + \sin\xi_0 \ s, \tag{20}$$

$$z(s) = z_0 + \cos\xi_0 \int_0^s \cos[\eta(s')]ds'.$$

III. SECONDARY ELECTRONS

In order to obtain approximately correct analytic formulas for electrons, let us consider a homogeneous medium, and suppose that the energy of the electron decreases linearly with depth (the "Harder law").[16] We let r_0 be the c.s.d.a. range,[17] so that

$$E(s) = S_T \ (r_0 - s), \tag{21}$$

with S_T a constant ($\cong$2 MeV/cm for water). The charge of an electron is

$$e = -q. \tag{22}$$

We may then perform the integration in Eq. (19) explicitly to obtain

$$\eta(s) = \eta_0 + \frac{qB}{S_T}\ln\left[\frac{\sqrt{\frac{S_T}{mc^2}r_0\left[\frac{S_T}{mc^2}r_0+2\right]}+\frac{S_T}{mc^2}r_0+1}{\sqrt{\frac{S_T}{mc^2}(r_0-s)\left[\frac{S_T}{mc^2}(r_0-s)+2\right]}+\frac{S_T}{mc^2}(r_0-s)+1}\right]. \tag{23}$$

If we then set

$$C(s) = \int_0^\tau \cos\left[\frac{qB}{S_T}\ln\left(\frac{\sqrt{\lambda(\lambda+2)}+\lambda+1}{\sqrt{\lambda(1-\tau')[\lambda(1-\tau')+2]}+\lambda(1-\tau')+1}\right)\right]d\tau', \tag{24}$$

and

$$S(s) = \int_0^\tau \sin\left[\frac{qB}{S_T}\ln\left(\frac{\sqrt{\lambda(\lambda+2)}+\lambda+1}{\sqrt{\lambda(1-\tau')[\lambda(1-\tau')+2]}+\lambda(1-\tau')+1}\right)\right]d\tau', \tag{25}$$

where

$$\lambda \equiv \frac{S_T r_0}{mc^2} = \frac{E_0}{mc^2}, \tag{26}$$

Eq. (20) becomes

$$x(s) = x_0 + r_0\cos\xi_0\left[\cos\eta_0 \ S\left(\frac{s}{r_0}\right) + \sin\eta_0 \ C\left(\frac{s}{r_0}\right)\right],$$

$$y(s) = y_0 + \sin\xi_0 \ s,$$

$$z(s) = z_0 + r_0\cos\xi_0\left[\cos\eta_0 \ C\left(\frac{s}{r_0}\right) - \sin\eta_0 \ S\left(\frac{s}{r_0}\right)\right]. \tag{27}$$

Let us look at some sample trajectories to see the effect of the magnetic field. We take the electron to start at $x_0 = y_0 = z_0 = 0$ with energy $E_0 = 10$ MeV, and to be initially directed along the z-axis, so that $\xi_0 = \eta_0 = 0$. According to Eq. (12.42) of Jackson,[15] we can use units of MeV and cm in Eq. (23) if we also set

$$q = 3.00 \tag{28}$$

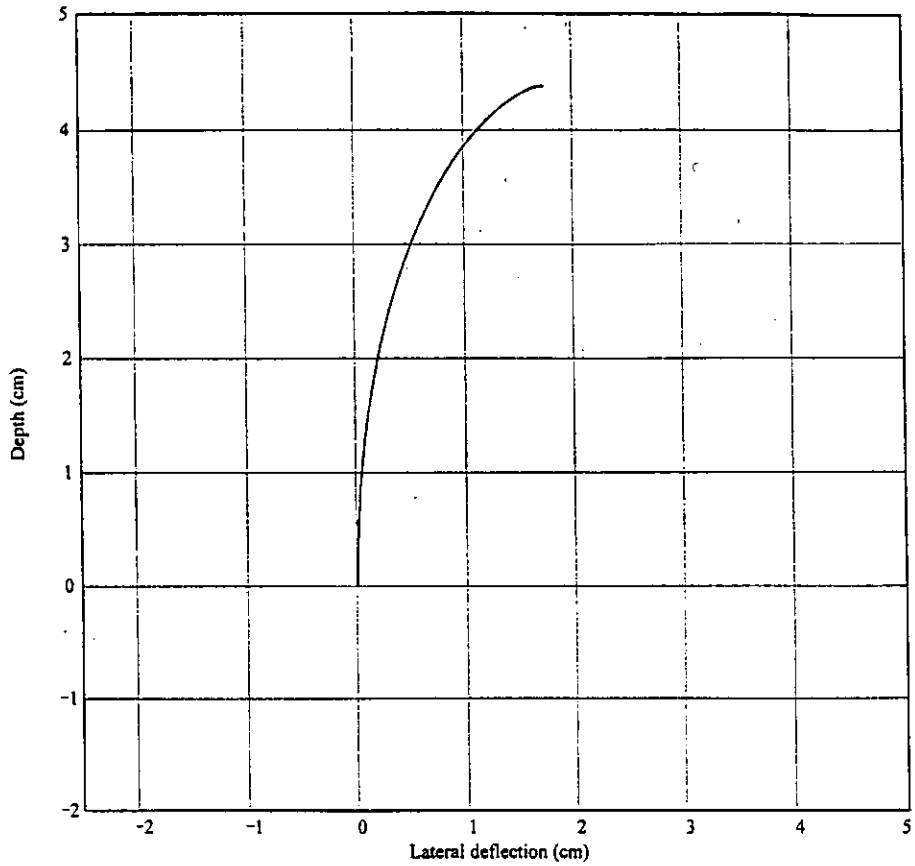


FIG. 1. Trajectory of a 10-MeV electron in water in the presence of a transverse magnetic field of strength $B=0.3$ T.

there, and give B in teslas (1 T$=10^4$ gauss). The electron stays in the x-z plane, of course, with trajectory given by

$$x(s)=r_0 S\left(\frac{s}{r_0}\right),$$
$$z(s)=r_0 C\left(\frac{s}{r_0}\right). \tag{29}$$

Figures 1–3 gives these trajectories for $B=0.3$, 2, and 5 T, and we see that the trajectories increasingly curl in on themselves for increasing B, as expected.

What we are interested in is the distribution of energy deposited in the medium, and we can easily calculate for a particular trajectory the amount of energy deposited between depths z and $z+\Delta z$ by dividing the path length into a great many segments and placing the dose deposited in that segment into the appropriate z-bin, as a histogram. In this computation we should exclude the energy carried off by bremsstrahlung photons and high-energy secondary electrons set in motion by the original electron, so that the dose deposited along a path segment of length Δs is

$$\Delta D=\frac{S_c}{\rho}\Delta s, \tag{30}$$

where S_c/ρ is the restricted mass collision stopping power (also assumed to be constant in these examples). This dose distribution for $B=2$ T, normalized to a total energy deposition of 1, is shown in Fig. 4. The fine structure of these distributions will of course get washed out for electrons set in motion at various depths in a photon beam; what is significant is the concentration of dose near the initial position of the electron, for large B.

In a real photon beam, the secondary electrons emerge from interactions with a wide range of angles with respect to the beam's direction and with a wide range of energies. In other articles[18,19] in a related series we have modeled the setting in motion, transport, and energy deposition of Compton electrons. First[18] we represented the photon beam by photons of several energies, and then, for each of these photon energies, we represented the Compton electrons emerging from a given interaction point with several "Compton cones" (actually the surfaces of cones) of electrons all of one energy and polar angle (and variable azimuthal angle). Another article[20] in that series provides a similar representation of the electrons and positrons created by the pair-production process. For the Compton electrons,[19] we developed a simple model for dealing with localized tissue inhomogeneities in the central portion of the photon beam, by considering planes of Compton cones. Thus we should next work out the dose deposited between z and $z+\Delta z$ by a Compton cone of electrons, for this will give the dose at a point from all the Compton electrons created at depth z_0 of a particular energy, in the central portion of a broad beam. (So far, however, our explicit formulas are applicable only to a homogeneous medium, and we are still assuming a uniform magnetic field in the y-direction, of course.)

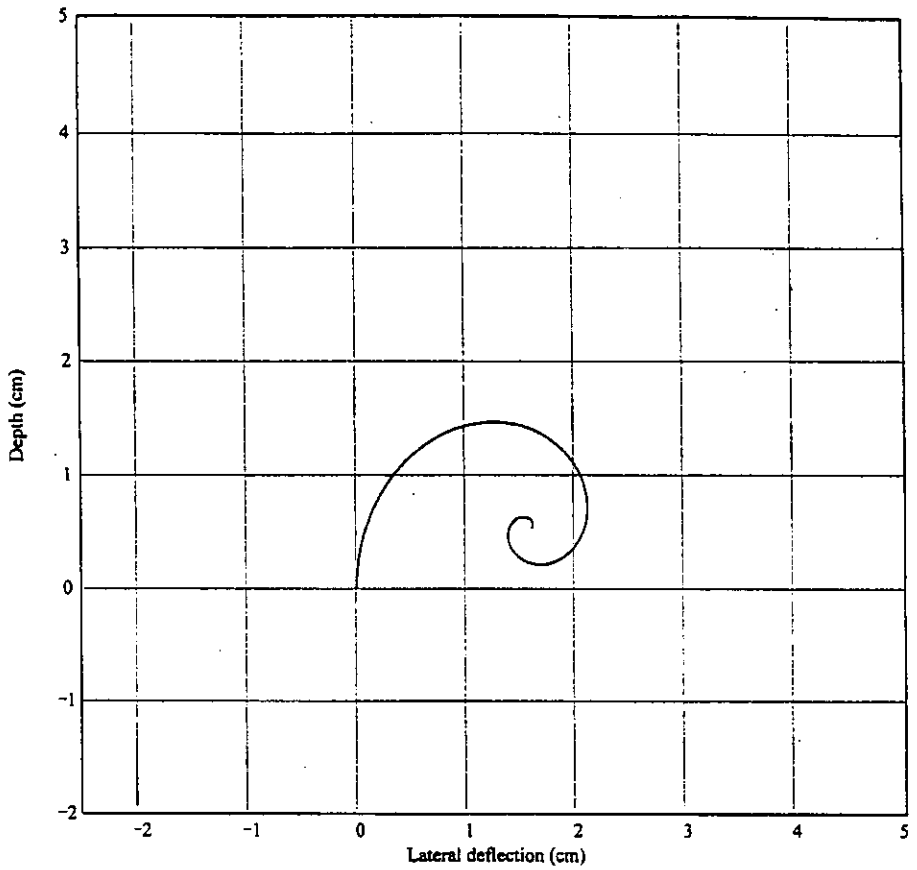


FIG. 2. Same as Fig. 1, but for $B = 2$ T.

We first change to a spherical coordinate system (θ,ϕ) with the z-axis (parallel to which the photon beam travels) as the polar direction. In this coordinate system, the initial direction of the electron is

$$\hat{v} = \sin\theta_0 \cos\phi_0 \hat{x} + \sin\theta_0 \sin\phi_0 \hat{y} + \cos\theta_0 \hat{z}. \tag{31}$$

Comparing this equation with Eq. (16) with $s = 0$, we see that Eq. (27) becomes

$$\begin{aligned} x(s) &= x_0 + r_0\left[\cos\theta_0 S\left(\frac{s}{r_0}\right) + \sin\theta_0 \cos\phi_0 C\left(\frac{s}{r_0}\right)\right], \\ y(s) &= y_0 + \sin\theta_0 \sin\phi_0 s, \\ z(s) &= z_0 + r_0\left[\cos\theta_0 C\left(\frac{s}{r_0}\right) - \sin\theta_0 \cos\phi_0 S\left(\frac{s}{r_0}\right)\right]. \end{aligned} \tag{32}$$

For 10 MeV electrons in water starting at $x_0 = y_0 = z_0 = 0$ with $\theta_0 = 15°$, Fig. 5 shows the projections onto the x-z plane of trajectories for $B = 2$ T, with five equally-spaced initial azimuthal angles ϕ_0: 18°, 54°, 90°, 126°, and 162°.

Actually, the dose contributions of the trajectories given by Eq. (32) should be averaged over all ϕ_0, and this is done in Fig. 6 for $B = 2$ T. We see that there is marked concentration of dose near the plane of the Compton cones, which helps to explain the dose enhancement observed by Li in his Monte Carlo calculations for a photon beam in a transverse magnetic field whose magnitude is constant in the transverse direction but steadily increasing in the longitudinal direction. In this configuration, the secondary electrons and positrons set in motion at depths smaller than that of the dose enhancement travel deeper because B is smaller there (at first). What one has to do is to sum up dose contributions from such planes of secondary electrons, and regions of less magnetic field strength B produce less concentration of dose. However, the variation of B with z leads to distortions of the trajectories which can be modeled starting from our new equation of motion, Eq. (13), and we are here providing only a qualitative explanation of this dose-enhancement phenomenon. We also note that there will be a corresponding region of dose reduction just distal to the dose-enhancement region, if the strength of the transverse magnetic field decreases with increasing depth there.

IV. MULTIPLE SCATTERING

The dominant process involved in the transport of electrons through a material is multiple scattering, causing the electrons to increasingly deviate laterally from their original line of motion with increasing penetration depth. In the absence of a strong magnetic field, one can derive the distribution function of the electrons through use of a small-angle approximation for polar angles (i.e., angles with respect to the original direction of motion). The Fermi–Eyges multiple-scattering theory[21,22] uses the usual first-order small-angle approximation,

$$\begin{aligned} \sin\theta &\cong \theta, \\ \cos\theta &\cong 1, \end{aligned} \tag{33}$$

 8690

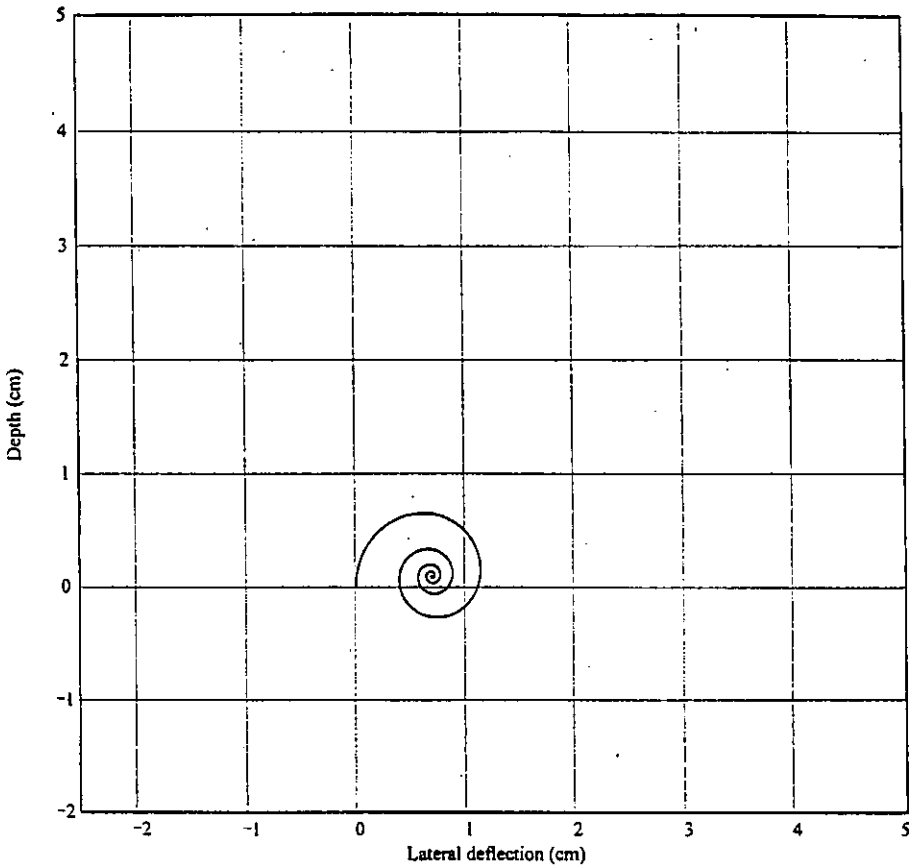


FIG. 3. Same as Fig. 1, but for $B = 5$ T.

and in our own theoretical work[23–26] on electron transport we have used a more accurate "second-order" small-angle approximation:[27]

$$\sin\theta \cong \theta - \frac{\theta^3}{6}, \quad \cos\theta \cong 1 - \frac{\theta^2}{2}. \tag{34}$$

Multiple-scattering theories based on the first-order small angle approximation have good accuracy to perhaps 50% of the range of the electrons, and use of the second-order small-angle approximation extends this to about 70% of the range. However, towards the end of the range the electrons spread out so much that they enter a diffusion régime which must be described by another multiple-scattering theory.

In the presence of a strong magnetic field $\vec{B}$, these small-angle approximations cannot be used directly, for the direction of the electrons quickly deviates significantly from their original direction of motion. This is seen in Figs. 1–3, in which, for 10-MeV electrons in water, we would expect to be able to apply a multiple-scattering theory based on a small-angle approximation for $B = 0.3$ T (except towards the end of the range of the particles), but certainly not for $B \geq 2$ T. It is

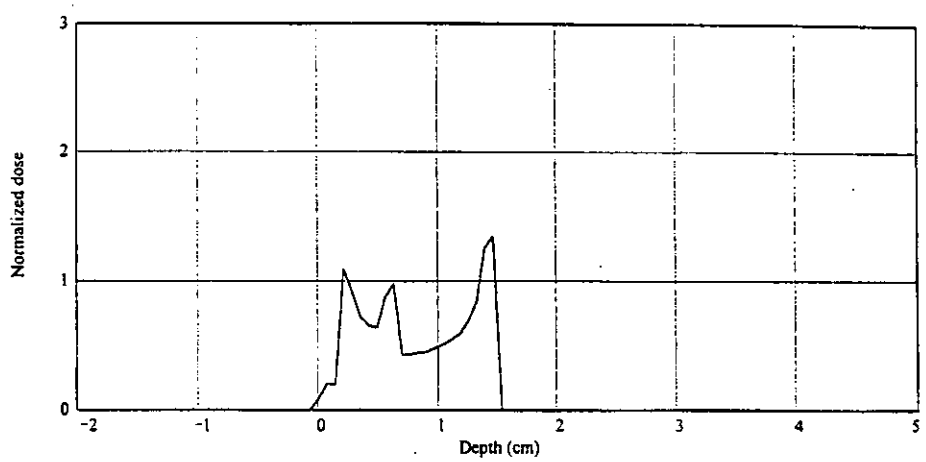


FIG. 4. Normalized dose deposited at depth z by a 10 MeV electron in water in the presence of a transverse magnetic field of strength $B = 2$ T.

 3691

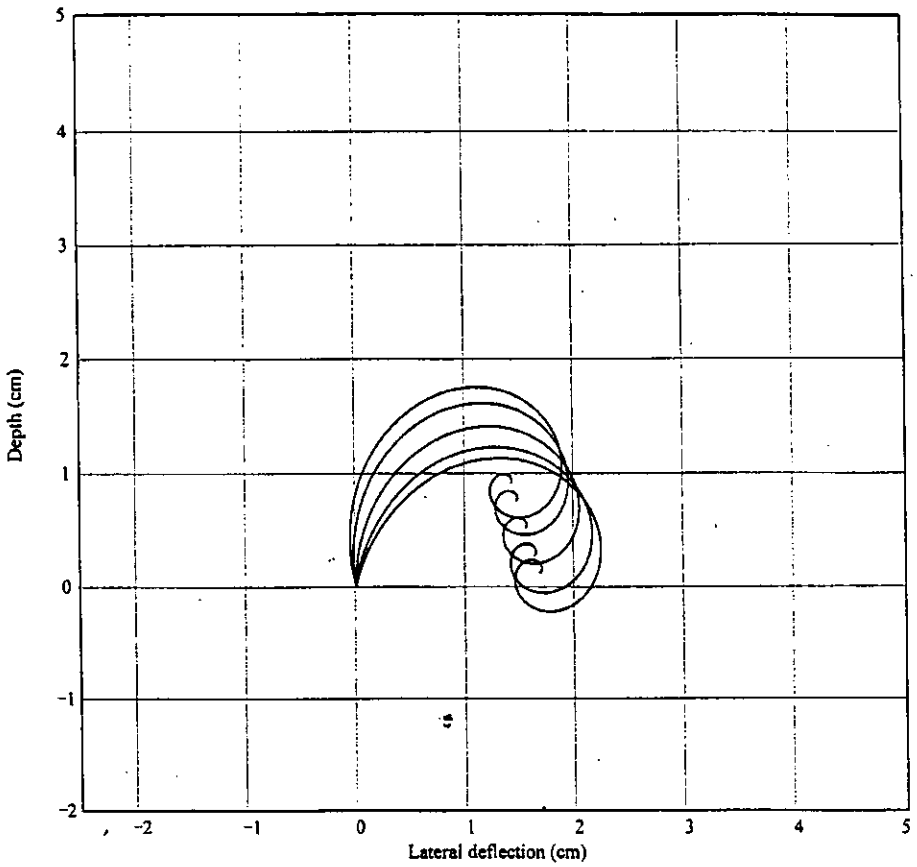


FIG. 5. Projections in the x-z plane of trajectories of 10-MeV electrons in water in the presence of a transverse magnetic field of strength $B=2$ T, with a polar angle θ_0 of 15° and azimuthal angles ϕ_0 of 18°, 54°, 90°, 126°, and 162°.

nonetheless possible to incorporate multiple scattering based on a small-angle approximation into our description of an electron's motion in a strong magnetic field, but this will require a radically new approach to such descriptions. We present this new theory in this section.

For the moment we don't worry about magnetic fields, and introduce the concept of a *typical scattered particle* (TSP). In two dimensions, we consider a ray of particles initially directed along the z-axis at $x=z=0$, and use the first-order small-angle approximation given by Eq. (33), so that the pathlength s of a particle can be approximated by the corresponding distance z along its original line of motion:

$$s \cong z. \tag{35}$$

The trajectory of a TSP is to be specified by

$$x=\alpha h(z) \cong \alpha h(s), \tag{36}$$

where α is the parameter of a family of such trajectories specified by the function h. Rather than the planar fluence $L(x,z)$ resulting from a multiple-scattering theory, we use the *density* of TSP's $\chi(\alpha)$ defined by

$$L(x,z)dx=\chi(\alpha)d\alpha, \tag{37}$$

so that, by Eq. (36),

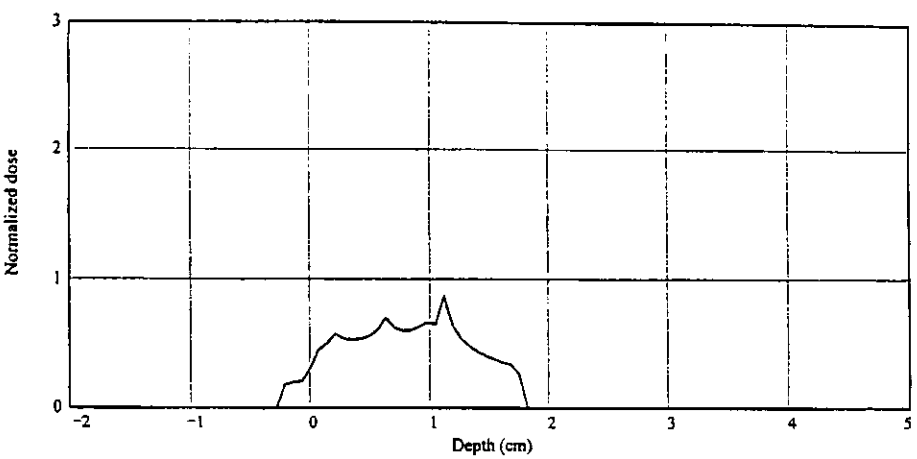


FIG. 6. Normalized dose deposited at depth z by a 10-MeV Compton cone in water in the presence of a transverse magnetic field of strength $B=2$ T.

$$\chi(\alpha)=L(\alpha h(z),z)h(z). \tag{38}$$

Evidently our underlying multiple-scattering theory must be such that there exists a function h such that the right-hand-side of Eq. (38) is independent of z, for any value of α. This is not true in general, although it is approximately true for our second-order multiple-scattering theory,[27] for our new multiple-scattering theory incorporating large-angle single scattering,[24] and for Molière multiple-scattering theory.[28] This is because such an h exists for any multiple-scattering theory for which

$$L(x,z)=\frac{e^{-x^2/\sigma^2(z)}}{\sqrt{\pi}\sigma(z)}, \tag{39}$$

where $\sigma(z)/\sqrt{2}$ is the projection onto the x-z plane of the rms lateral deviation of the electrons at depth z, for we may then choose

$$h(z)=\sqrt{\sigma(z)}, \tag{40}$$

and have

$$\chi(\alpha)=\frac{e^{-\alpha^2}}{\sqrt{\pi}}. \tag{41}$$

Equation (39) holds exactly for the Fermi–Eyges theory, with

$$\sigma(z)=A_2(z)\equiv\int_0^z T(\xi)(z-\xi)^2\,d\xi, \tag{42}$$

where T is the scattering power.[17]

The concept of typical scattered particles is similar to that of peak representative path, which was first put forward by Perry and Holt[29] and used by Yu *et al.*[30] in their multiray model for electron dose calculation. The peak representative path is found by looking at all the particles which actually arrive at a given point (x,z) in Fig. 7; at any depth z', the coordinate x' of the peak representative path is the maximum of the (Gaussian) distribution of these particles at that depth. The physical importance of the representative path method is that, at any given depth z', most of the particles lie close to x', with a standard deviation from x' which can be calculated; therefore one can approximately calculate the effect of localized tissue inhomogeneities by (mathematically) following these representative paths through the inhomogeneity, as we have illustrated elsewhere.[31,32] Perry and Holt[29] considered only the approximation of no energy loss of the particles in their formulas for peak representative path and standard deviation from this path, but we have used Fermi–Eyges theory to determine these quantities for energy loss; for the peak representative path, x' is given by

$$x'=\frac{A_2(z')+(z-z')A_1(z')}{A_2(z)}x, \tag{43}$$

where

$$A_n(z)=\int_0^z T(\xi)(z-\xi)^n\,d\xi, \tag{44}$$

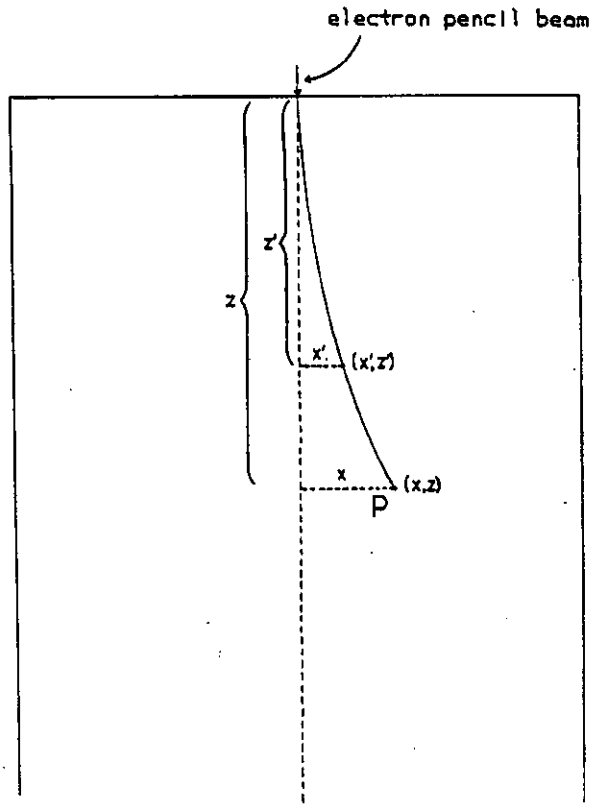


FIG. 7. Trajectory of typical scattered particle or of peak representative path.

as in Eq. (42).

Typical scattered particles, on the other hand, do not have an immediate physical meaning, for they are just a mathematical construct. However, we shall be able to claim that they are useful for dealing with localized inhomogeneities, for example, if the path of a TSP lies very close to the corresponding peak representative path. We use the Fermi–Eyges theory to specify the TSP path, for which Eqs. (36), (40), and (42) imply

$$x'=\sqrt{\frac{\sigma(z')}{\sigma(z)}}x=\sqrt{\frac{A_2(z')}{A_2(z)}}x. \tag{45}$$

From Eqs. (43) and (45), we evidently must see how close the functions

$$f(z',z)\equiv\sqrt{\frac{A_2(z')}{A_2(z)}} \tag{46}$$

and

$$g(z',z)\equiv\frac{A_2(z')+(z-z')A_1(z')}{A_2(z)} \tag{47}$$

are to each other. Defining

$$h(z',z)=f(z',z)-g(z',z), \tag{48}$$

and using

$$\frac{dA_{n+1}(z)}{dz}=(n+1)A_n(z), \tag{49}$$

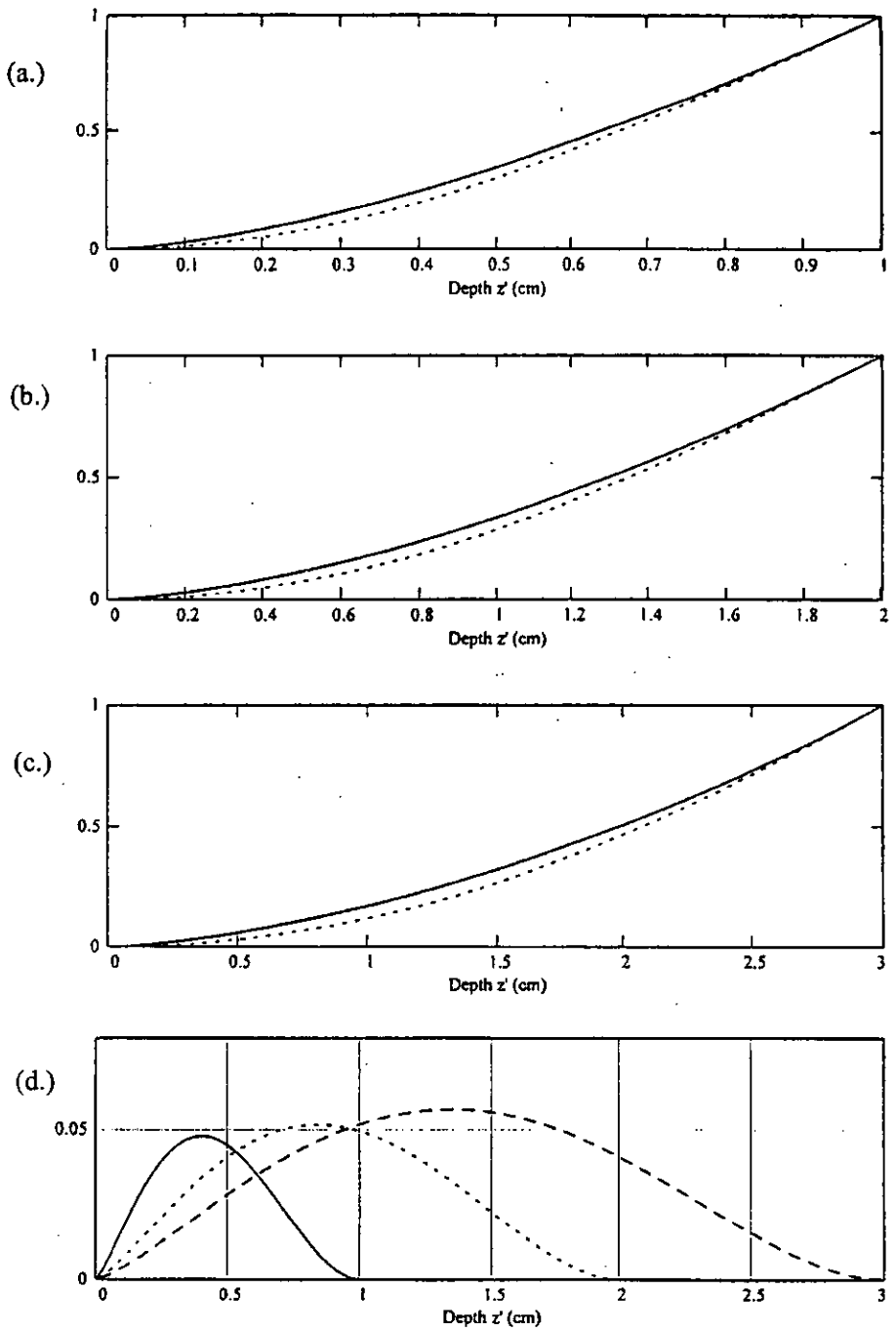


FIG. 8. Comparisons of the functions f given by Eq. (46) (for typical scattered particles) and g given by Eq. (47) (for peak representative paths), for 10-MeV electrons in water. (a) f (solid curve) and g (dashed curve) for $z=1$ cm. (b) Same, for $z=2$ cm. (c) Same, for $z=3$ cm. (d) Difference $f-g$ for $z=1$ cm (solid curve), $z=2$ cm (short-dash curve), and $z=3$ cm (long-dash curve).

as well as

$$A_n(z)\cong\frac{T|_{z=0}\ z^{n+1}}{n+1}, \tag{50}$$

for small z, we readily find that all of the following quantities are zero:

$$h(0,z)=h(z,z)$$
$$=\frac{d}{dz'}[h(z',z)]|_{z'=0}$$
$$=\frac{d}{dz'}[h(z',z)]|_{z'=z}=0. \tag{51}$$

Therefore we fully expect h to be small compared to 1 (the value of f and g at $z'=z$), and this is confirmed in Fig. 8 for 10-MeV electrons in water: the maximum deviation of f from g for the three depths z considered (1, 2, and 3 cm) is seen to be only 0.06. Thus our TSP's are physically meaningful.

Let us therefore integrate our TSP's into our equation of motion, Eq. (13), by first setting $\vec{B}=\vec{0}$ and asking whether there exists a function g for which the desired family of TSP's is produced by the following equation of motion:

$$\frac{d\vec{p}}{dt}=-S_T(E(s))\frac{\vec{v}}{v}-\frac{\alpha g(s)}{c}\vec{v}\times\hat{v}_\perp, \tag{52}$$

 7694

where $\hat{v}_\perp$ is a (any) unit vector perpendicular to $\hat{v}_0$, the initial direction of the particle. Comparing Eqs. (13) and (52), we see that for a particle starting at the origin and initially directed along the z-axis, and with the choice $\hat{v}_\perp = \hat{y}$, we shall obtain the same trajectory as that given by Eqs. (19), (20) with $x_0 = y_0 = z_0 = \xi_0 = \eta_0 = 0$ and

$$-eB = qB \rightarrow \alpha g(s) \tag{53}$$

[and placed inside the integral in Eq. (19)], i.e.,

$$x(s) = \int_0^s \sin\left(\alpha \int_0^{s'} \frac{g(s'')ds''}{\sqrt{E(s'')[E(s'')+2mc^2]}} \right) ds',$$

$$y(s) = 0, \tag{54}$$

$$z(s) = \int_0^s \cos\left(\alpha \int_0^{s'} \frac{g(s'')ds''}{\sqrt{E(s'')[E(s'')+2mc^2]}} \right) ds'.$$

We want this trajectory to be the same as that given by Eq. (36), and we see that this is possible to do under the first-order small-angle approximation, Eq. (33), if

$$\int_0^s \left[\int_0^{s'} \frac{g(s'')ds''}{\sqrt{E(s'')[E(s'')+2mc^2]}} \right] ds' = h(s). \tag{55}$$

Thus in general we must choose

$$g(s) = \sqrt{E(s')[E(s')+2mc^2]} \frac{d^2h(s)}{ds^2}, \tag{56}$$

and for h given by Eq. (40) we have

$$g(s) = \frac{\sqrt{E(s)[E(s)+2mc^2]}}{4[\sigma(s)]^{3/2}} \left\{ 2\sigma(s) \frac{d^2\sigma(s)}{ds^2} - \left[\frac{d\sigma(s)}{ds} \right]^2 \right\}. \tag{57}$$

For the Fermi–Eyges theory, Eqs. (42) and (49) imply

$$g(s) = \sqrt{E(s)[E(s)+2mc^2]} \frac{A_0(s)A_2(s) - [A_1(s)]^2}{[A_2(s)]^{3/2}}, \tag{58}$$

and the A_n are given explicitly by Eqs. (142)–(144) of Ref. 19; for $E(s)$ either Eq. (21) or the more accurate Eq. (130) of Ref. 19 can be used.

By combining Eqs. (13) and (52), we thus obtain the equation of motion of our TSP's in an arbitrary magnetic field $\vec{B}(x,y,z)$:

$$\frac{d\vec{p}}{dt} = -S_T(E(s)) \frac{\vec{v}}{v} - \frac{\alpha g(s)}{c} \vec{v} \times \hat{v}_\perp + \frac{e}{c} \vec{v} \times \vec{B}. \tag{59}$$

Let us again consider a uniform magnetic field in the y-direction as specified by Eq. (15), and suppose that the particle has initial position (x_0, y_0, z_0) and initial direction

$$\hat{v}_0 = \cos\xi_0 \sin\eta_0 \ \hat{x} + \sin\xi_0 \ \hat{y} + \cos\xi_0 \cos\eta_0 \ \hat{z}. \tag{60}$$

Let us choose $\hat{v}_1$ to be a vector in the x-z plane perpendicular to $\hat{v}_0$,

$$\hat{v}_1 \equiv \cos\eta_0 \ \hat{x} - \sin\eta_0 \ \hat{z}, \tag{61}$$

and let us choose $\hat{v}_2$ to be $\hat{v}_0 \times \hat{v}_1$:

$$\hat{v}_2 \equiv -\sin\xi_0 \sin\eta_0 \ \hat{x} + \cos\xi_0 \ \hat{y} - \sin\xi_0 \cos\eta_0 \ \hat{z}. \tag{62}$$

Then, for an arbitrary angle ω, we may define

$$\hat{v}_\perp \equiv \cos\omega \ \hat{v}_1 + \sin\omega \ \hat{v}_2, \tag{63}$$

i.e.,

$$\hat{v}_\perp \equiv (\cos\eta_0 \cos\omega - \sin\xi_0 \sin\eta_0 \sin\omega)\hat{x} + \cos\xi_0 \sin\omega \ \hat{y} - (\sin\eta_0 \cos\omega + \sin\xi_0 \cos\eta_0 \sin\omega)\hat{z}. \tag{64}$$

As in deriving Eq. (14), from Eq. (59) we deduce

$$\sqrt{E(s)[E(s)+2mc^2]} \frac{d\hat{v}}{ds} = -\alpha g(s) \hat{v} \times \hat{v}_\perp + eB\hat{v} \times \hat{y}. \tag{65}$$

Again using Eq. (16) for $\hat{v}(s)$, we readily find that Eq. (65) is equivalent to the following two equations:

$$\sqrt{E(s)[E(s)+2mc^2]} \frac{d\xi}{ds} = -\alpha g(s) \begin{bmatrix} (\sin\eta_0 \cos\omega + \sin\xi_0 \cos\eta_0 \sin\omega)\sin\eta + \\ (\cos\eta_0 \cos\omega - \sin\xi_0 \sin\eta_0 \sin\omega)\cos\eta \end{bmatrix}, \tag{66}$$

and

$$\sqrt{E(s)[E(s)+2mc^2]} \frac{d\eta}{ds} = -eB + \alpha g(s) \times \left\{ \tan\xi \begin{bmatrix} (\sin\xi_0 \sin\eta_0 \sin\omega - \cos\eta_0 \cos\omega)\sin\eta + \\ (\sin\eta_0 \cos\omega + \sin\xi_0 \cos\eta_0 \sin\omega)\cos\eta \end{bmatrix} + \cos\xi_0 \sin\omega \right\}. \tag{67}$$

For $B \neq 0$, it appears that Eqs. (66), (67) do not admit an analytic solution, except for the special case discussed below. [If $B = 0$, we would be able to solve Eqs. (66), (67) by transforming to a different coordinate system, and we would get the TSP trajectories associated with straight-line motion for $\alpha = 0$.] Nonetheless, these equations determine a unique solution for given initial values ξ_0 and η_0 of ξ and η, and a particular choice of ω, and this solution can be calculated through a straightforward iterative procedure, or by a standard mathematical software package such as MATHCAD. Thus we have various TSP trajectories associated with values of α and ω, and with them we can calculate quantities such as dose, as was done in the previous section. Suppose that the contribution to some quantity Q is $\tilde{Q}(\alpha,\omega)$, for a particular TSP trajectory. Then we would calculate Q as

$$Q = \int_0^{2\pi} d\omega \int_0^\infty \alpha\chi(\alpha)\tilde{Q}(\alpha,\omega)d\alpha, \tag{68}$$

after generalizing Eq. (37) determining χ to three dimensions. For example, in place of Eq. (41) we would use

$$\chi(\alpha) = \frac{e^{-\alpha^2}}{\pi}. \tag{69}$$

Furthermore, in dealing with Compton cones we would make in Eqs. (66) and (67) the substitutions indicated by equating Eqs. (31) and (60) to each other, i.e.,

$$\sin\xi_0 \sin\eta_0 = \sin\theta_0 \cos\phi_0,$$
$$\cos\xi_0 = \sin\theta_0 \sin\phi_0, \tag{70}$$
$$\sin\xi_0 \cos\eta_0 = \cos\theta_0;$$

then we would average over ϕ_0, as was done in obtaining Fig. 6. (The effect for that example would be to largely blur out the fine structure of the dose distribution, as well as to somewhat broaden the dose distribution.)

Finally, we present what is apparently the only analytic solution of Eqs. (66), (67) for $B \neq 0$. We shall set the right-hand-side of Eq. (66) to zero, and this will require $\xi_0 = 0$ and $\omega = \pi/2$. This means that the initial direction of the particle is in the x-z plane, and we are choosing $\hat{v}_\perp = \hat{y}$. Then Eq. (67) becomes

$$\frac{d\eta}{ds} = -\frac{eB}{\sqrt{E(s)[E(s)+2mc^2]}} + \frac{\alpha g(s)}{\sqrt{E(s)[E(s)+2mc^2]}}, \tag{71}$$

and with the use of Eq. (55) we conclude

$$\eta(s) = \eta_0 - eB\int_0^s \frac{ds'}{\sqrt{E(s')[E(s')+2mc^2]}} + \alpha\frac{dh(s)}{ds}, \tag{72}$$

for insertion into Eqs. (20). For the Fermi–Eyges theory, we have

$$\frac{dh(s)}{ds} = \frac{1}{2\sqrt{\sigma(s)}}\frac{d\sigma(s)}{ds} = \frac{A_1(s)}{\sqrt{A_2(s)}}. \tag{73}$$

V. CONCLUSION

We have developed a new equation of motion for the transport of charged particles in an arbitrary magnetic field. This equation, Eq. (59), includes the loss of energy according to the continuous slowing-down approximation, and our new concept of typical scattered particles allows it to incorporate the multiple-scattering process. The formulas which we have arrived at are particularly applicable to the transport of, and deposition of energy by, Compton electrons and pair-production electrons and positrons generated within a medium by a photon beam, and we have shown qualitatively how a large dose enhancement can occur within the treatment volume.

One major difficulty which we have mentioned at the beginning of Sec. IV is the problem of describing the diffusion regime near the end of the particles' range with an analytic multiple-scattering theory. This problem is compounded by the presence of occasional large-angle elastic scattering, although we have previously developed a multiple-scattering theory[24] which does take this process into account. We shall have to examine the extent to which introducing magnetic fields into a small-angle formalism such as the one presented here compounds this end-of-range problem.

The present work but lays the theoretical basis for modeling the dose in a photon beam in the presence of a strong magnetic field. To carry out such study systematically, it will be necessary to use a realistic magnetic field, and in a companion article[33] we have accordingly provided a model field which can be expected to exhibit the desired large localized dose enhancement and corresponding dose reduction just distal to the dose-enhancement region; in that article we have carried out Monte Carlo calculations to examine this effect. Further study will have to include a considerable amount of Monte Carlo calculation to gain additional insights and to confirm the accuracy of particular theoretical results. It will be most important to investigate the influence of the magnetic field on the dose distribution in the vicinity of regions where there is not electronic equilibrium, such as distal to localized tissue inhomogeneities and in the penumbral region. Another very important, and mathematically quite complicated, question to pursue is the possibility of choosing the magnetic field in such a way as to produce a desired dose distribution. When and if it becomes technologically possible to produce such magnetic fields, their use may be expected to provide significant improvement in radiation treatment for certain types of cancer.

ACKNOWLEDGMENTS

The author is indebted to Allen Li, James C. H. Chu, and Leonard Reiffel for providing the impetus for carrying out the present theoretical study, as well as for ongoing discussions concerning it. These investigations were supported by a grant from the National Cancer Institute (No. 2 R01 CA41490).

[a]Electronic mail: dave@lanzl.com

[1] C. C. Shih, "High energy electron radiotherapy in a magnetic field," Med. Phys. **2**, 9–13 (1975).
[2] D. P. Whitmire, D. L. Bernard, and M. D. Peterson, "Magnetic enhancement of electron dose distribution in radiation therapy," Proc. 4th Conf. Applications of Small Accelerators (1976), published in IEEE Trans. Nucl. Sci. (1997).
[3] D. P. Whitmire, D. L. Bernard, M. D. Peterson, and J. A. Purdy, "Magnetic enhancement of electron dose distribution in a phantom," Med. Phys. **4**, 127–131 (1977).
[4] D. P. Whitmire, D. L. Bernard, and M. D. Peterson, "Magnetic modification of the electron-dose distribution in tissue and lung phantoms," Med. Phys. **5**, 409–417 (1978).
[5] B. R. Paliwal, A. L. Wiley, Jr., B. W. Wessels, and M. C. Choi, "Magnetic field modification of electron-beam dose distributions in inhomogeneous media," Med. Phys. **5**, 404–408 (1978).
[6] B. R. Paliwal, B. R. Thomadsen, and A. L. Wiley, Jr., "Magnetic modification of electron beam dose distributions," Acta Radiol.: Oncol. **18**, 57–64 (1979).
[7] B. R. Paliwal, C. Pocheng, A. J. Greenberg, and A. L. Wiley, "A technique to improve the homogeneity of electron dose distribution in chest wall irradiation," Int. J. Radiat. Oncol., Biol., Phys. **5**, 1889–1892 (1979).
[8] R. Nath, "Physical and biological aspects of magnetic enhancement of high-energy electron dose distributions," in *Proceedings of the Symposium on Electron Dosimetry and Arc Therapy, Madison, Wisconsin, 1981* (American Institute of Physics, New York, 1982).
[9] M. S. Weinhous, R. Nath, and R. J. Schulz, "Enhancement of electron beam dose distributions by longitudinal magnetic fields: Monte Carlo simulations and magnet system optimization," Med. Phys. **12**, 598–603 (1985).

 3696

[10] A. F. Bielajew, "The effect of strong longitudinal magnetic fields on dose deposition from electron and photon beams," Med. Phys. **20**, 1171–1179 (1993).
[11] W. R. Nelson, H. Hirayama, and D. W. O. Rogers, "The EGS4 code system," Stanford Linear Accelerator Center Report 265 (Stanford University, Stanford, California, 1985).
[12] E. Nardi and G. Barnea, "Electron beam therapy with transverse magnetic fields," Med. Phys. **26**, 967–973 (1999).
[13] L. Reiffel, "Method and apparatus to control photon beam dose enhancements," U.S. Patent No. 5,974,112 issued October 26, 1999. See also L. Reiffel, X. A. Li, J. Chu, R. W. Wheatley, S. Naqvi, R. Pillsbury, and A. Saxena, "Control of dose profiles along therapeutic photon beams by topically applied transverse magnetic fields" (unpublished).
[14] J. C. H. Chu, L. Reiffel, S. Naqvi, X. A. Li, S.-J. Ye, and A. Saxena, "The use of magnetic fields to improve photon dose distributions for radiation therapy—a possible approach to 'poor man's proton' beam properties," World Congress on Medical Physics and Biomedical Engineering, Chicago, Illinois, July 23–28, 2000 (submitted).
[15] J. D. Jackson, *Classical Electrodynamics*, 2nd ed. (Wiley, New York, 1975).
[16] D. Harder, in *Symposium on High Energy Electrons*, edited by A. Zuppinger and G. Poretti (Springer, Berlin, 1965), pp. 26–33.
[17] ICRU, International Commission on Radiation Units and Measurements Report No. 35: "Radiation dosimetry: Electron beams with energies between 1 and 50 MeV" (Bethesda, MD, 1984).
[18] D. Jette, "Photon dose calculation based on electron multiple-scattering theory: Practical representation of dose and particle transport integrals," Med. Phys. **26**, 924–930 (1999).
[19] D. Jette, "Photon dose calculation based on electron multiple-scattering theory: Dose deposited by Compton electrons," Med. Phys. (submitted).
[20] L. Wang and D. Jette, "Photon dose calculation based on electron multiple-scattering theory: Primary dose deposition kernels," Med. Phys. **26**, 1454–1465 (1999).
[21] D. Jette, A. Pagnamenta, L. H. Lanzl, and M. Rozenfeld, "The application of multiple scattering theory to therapeutic electron dosimetry," Med. Phys. **10**, 352–355 (1983).
[22] D. Jette, "Electron dose calculation using multiple-scattering theory. A Gaussian multiple scattering theory," Med. Phys. **15**, 123–137 (1988); Med. Phys. **16**, 920(E) (1989).
[23] D. Jette, "Electron dose calculation using multiple-scattering theory: Localized inhomogeneities—A new theory," Med. Phys. **18**, 123–132 (1991).
[24] D. Jette, "Electron dose calculation using multiple-scattering theory: A new theory of multiple scattering," Med. Phys. **23**, 459–477 (1996). Med. Phys. **24**, 471(E) (1997).
[25] D. Jette and S. Walker, "Electron dose calculation using multiple-scattering theory: Energy distribution due to multiple scattering," Med. Phys. **24**, 383–400 (1997).
[26] D. Jette, "Electron dose calculation using multiple-scattering theory: Initial angular distribution," Med. Phys. (submitted).
[27] D. Jette and A. Bielajew, "Electron dose calculation using multiple scattering theory: Second-order multiple scattering theory," Med. Phys. **16**, 698–711 (1989).
[28] H. A. Bethe, "Molière's theory of multiple scattering," Phys. Rev. **89**, 1256–1266 (1953).
[29] D. J. Perry and J. G. Holt, "A model for calculating the effects of small inhomogeneities on electron beam dose distributions," Med. Phys. **7**, 207–215 (1980).
[30] C. X. Yu, W. S. Ge, and J. W. Wong, "A multiray model for calculating electron pencil beam distribution," Med. Phys. **15**, 662–671 (1988).
[31] D. Jette, "The problem of electron dose calculation: II. Inhomogeneities and beam shaping," J. Am. Assn. Med. Dosim. **9**, 12–17 (1984).
[32] D. Jette, "Electron beam dose calculations," in *Radiation Therapy Physics*, edited by A. R. Smith (Springer-Verlag, Berlin, 1995), pp. 95–121.
[33] D. Jette, "Magnetic fields with photon beams: Monte Carlo calculations for a model magnetic field," Med. Phys. (submitted).

 3697

Volume-based geometric modeling for radiation transport calculations

Zuofeng Li and Jeffrey F. Williamson[a)]
Radiation Oncology Center, Mallinckrodt Institute of Radiology, Washington University School of Medicine, St. Louis, Missouri 63110

(Received 7 May 1991; accepted for publication 7 February 1992)

Accurate theoretical characterization of radiation fields is a valuable tool in the design of complex systems, such as linac heads and intracavitary applicators, and for generation of basic dose calculation data that is inaccessible to experimental measurement. Both Monte Carlo and deterministic solutions to such problems require a system for accurately modeling complex 3-D geometries that supports ray tracing, point and segment classification, and 2-D graphical representation. Previous combinatorial approaches to solid modeling, which involve describing complex structures as set-theoretic combinations of simple objects, are limited in their ease of use and place unrealistic constraints on the geometric relations between objects such as excluding common boundaries. A new approach to volume-based solid modeling has been developed which is based upon topologically consistent definitions of boundary, interior, and exterior of a region. From these definitions, FORTRAN union, intersection, and difference routines have been developed that allow involuted and deeply nested structures to be described as set-theoretic combinations of ellipsoids, elliptic cylinders, prisms, cones, and planes that accommodate shared boundaries. Line segments between adjacent intersections on a trajectory are assigned to the appropriate region by a novel sorting algorithm that generalizes upon Siddon's approach. Two 2-D graphic display tools are developed to help the debugging of a given geometric model. In this paper, the mathematical basis of our system is described, it is contrasted to other approaches, and examples are discussed.

Key words: Geometric modeling, Monte Carlo simulation, radiation transport calculation, point-set topology

I. INTRODUCTION

A promising development in radiation dosimetry[1] is the use of radiation transport theory to solve dosimetric problems of clinical interest in radiation oncology, diagnostic imaging, and nuclear medicine. In contrast to the widely used empirical methods to estimate absorbed dose in the patient, transport theory begins with a mathematical description of the basic photon scattering and absorption processes underlying the transport and deposition of energy in matter along with a geometric description of the absorbing media and radiation sources in and around the patient. A central result[2] of transport theory is that, in principle, the Boltzmann transport equation completely characterizes the distribution of ionizing radiation, and related quantities such as adsorbed dose, within such a system. Methods[2] for numerically solving the transport equation include Monte Carlo simulation, the method of discrete ordinates, integral transport methods, and approximate methods such as the differential-pencil beam and dose-spread-array approaches. Of these, only Monte Carlo simulation has achieved the level of mathematical rigor, generality, and numerical robustness to qualify as a dosimetry tool, complementary to direct measurement, for estimating absorbed dose in complex three-dimensional (3-D) geometries.[2]

A general and robust method of modeling complex anatomic and artificially introduced structures, which supports ray tracing, point classification, and segment classification, is essential to the successful application of numerical transport methods to clinically relevant problems. Two approaches have been developed to model such structures. Three dimensional and transaxial imaging modalities, such as CT and MRI, are now widely used to quantitatively define patient anatomy. This approach yields a description of patient geometry consisting of a 3-D array of small volume elements or "voxels," each of which is assigned material properties (e.g., photon cross sections) to complete the description. Specialized ray-tracing algorithms, which exploit the geometrically regular arrangement of voxels, have been developed that allow theoretical dose estimates to be made by efficiently tracking photon trajectories through this materially heterogeneous 3-D array of voxels.[3] When manufactured devices are introduced to control and modify the dose distribution, another class of geometric modeling problems arises. Examples of such devices include linear accelerators equipped with beam modifying filters and shielded gynecological applicators containing sealed radioactive sources. These devices are most easily described as set-theoretic combinations (unions, intersections, etc.) of elliptic cones and cylinders, angled planes, cuboids, and other volumes that are easily describable in terms of analytic geometry. We have developed a general and robust combinatorial geometric modeling system to support dosimetric characterization of such systems using Monte Carlo simulation and other radiation transport calculation methods.

Combinatorial modeling systems have been used in numerical transport calculations for at least three decades.[4–7] In the public-domain transport-code packages, however, a

 0094-2405/92/30667-11$01.20

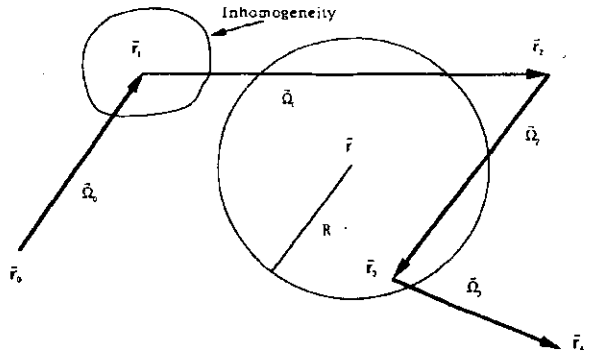


FIG. 1. A two-dimensional representation of a typical photon history. The vectors Ω_j denotes the trajectory of the photon leaving its jth collision at location $\mathbf{r}$.

pathological condition known as the touching boundary problem is excluded, leading to unnecessary complications in the application of these systems. Our geometric modeling code avoids this problem by adhering to a rigorous definition of the boundary of geometric objects and allows touching boundaries.

In addition, we have developed a novel segment classification algorithm which orders segments of a ray between intersections with the geometric regions along with the media containing each segment.

To facilitate the verification and error-checking of geometric models developed using our geometric modeling package, two graphic tools are developed. They have proven to be of great help in debugging complex geometric models.

II. METHOD AND THEORY

A. Problem definition and algorithm organization

Monte Carlo simulation of the transport equation consists of constructing, by means of a digital computer, a randomly selected set of particle trajectories from the set of all possible trajectories. The quantity of interest is then averaged over this small set giving a statistical estimate of the desired quantity.[1]

A two-dimensional representation of a typical photon history is shown in Fig. 1. The vector $\beta_j = (\mathbf{r}_j, \Omega_j, E_j)$ represents the state of the photon leaving its jth intersection or collision at position $\mathbf{r}_j$ with energy E_j and direction Ω_j. Each history α can be represented by a sequence of states of the form $\alpha = (\beta_0, \beta_1, \ldots)$. As interactions between the photon and the medium are limited to the discrete points $(\mathbf{r}_0, \ldots, \mathbf{r}_j \ldots)$, the trajectories connecting these collision sites are straight lines. The boundaries define 3-D regions in space that have well-defined radiological properties, e.g., cross-section tables, associated with them.

Fundamental to any transport algorithm is the need to evaluate the attenuation of photons between any two points, $\mathbf{r}_1$ and $\mathbf{r}_2$, by the regions traversed by the particle flight path. One must find the intersections, $\mathbf{e}_j = (\mathbf{r}_1 + d_j \Omega_1)$, of the flight path, with any region boundaries it encounters. Then the particle fluence at $\mathbf{r}_2$ can be written as

$$\Phi(\mathbf{r}_2) = \frac{\Phi(\mathbf{r}_1) \cdot e^{-\Sigma \mu_j l_j}}{|\mathbf{r}_2 - \mathbf{r}_1|^2},$$

where $l_j = (d_{j+1} - d_j)$ is the length of the line segment formed by adjacent intersections $\mathbf{e}_{j+1}$ and $\mathbf{e}_j$.

This process consists of four well-defined components:

(1) Given any point in 3-D space, the region containing this point must be found, i.e., the modeling system must support topological point classification;

(2) all intersections of a given ray with the boundaries of all regions characterized by different radiological properties must be found;

(3) the region containing each ray-path segment (and the medium traversed) l_j defined by adjacent region intersections $\mathbf{e}_{j+1}$ and $\mathbf{e}_j$ must be identified, i.e., the modeling system must support topological segment classification;

(4) for debugging of complex geometry input files, a method of graphically displaying the intersection of a complex geometric structure with an arbitrary plane is needed.

Component (4) is, of course, not a mathematically essential element of geometric modeling. Quantitative description of complex structures by human users is not a simple or easy mental activity. We have found graphical-display aids to be essential in construction, debugging, and verification of complex geometric models.

Our study of volume-based geometric modeling concentrates on systems of adsorbing media consisting of a finite number of regions $\{R_1, \ldots, R_n\}$. Each region R_i except the outermost one, must be wholly contained in some other region $R_j = \mathrm{EXT}(R_i)$ where $y = \mathrm{EXT}(x)$ can be interpreted as "volume y is the smallest region not equal to x which completely contains x." In addition, the regions immediately contained in each region $\mathrm{INT}(R_i) = \{R_j, R_k, R_m, \text{etc.}\}$ are listed. Region interiors must be either disjoint relative to one another or wholly contained in one another although they are allowed to share common boundaries. In addition, each region is assigned a density and a photon or electron cross-section table. We also require that the regions have the mathematical property of being orientable, i.e., for each region a continuous function $F(x, y, z)$ exists that describes the region's surface such that $F(x, y, z)$ is larger than zero on one side of the region surface and less than zero on the other side of the surface. This is required to implement topological segment classification.

Each region R_i is assumed to consist of a topological combination of a finite number of geometric objects or components, $\{O_{i1}, \ldots, O_{ij}, \ldots, O_{in_i}\}$. Letting $\cap$, $\cup$, and $\sim$ denote intersection, union, and complement, respectively, then each region is represented as an ordered sequence of form such as $R_i = [(O_{i1} \cup O_{i2}) \sim \cap O_{i3}] \cap O_{i4}$.

Each component set O_{ij} is defined by specifying its geometric type, origin, a set of coefficients specifying its size and orientation, and a set-theoretic operator $C_j \in \{\cap, \cup, \sim\cap, \text{or} \sim\cup\}$. Currently implemented geometric types include plane, ellipsoid, elliptic right cylinder, elliptic cone, rectangular parallelepiped, and unbounded medium. The components O_{ij} are required to be convex

and to contain open interiors in E^3 (three-dimensional real space with the usual Euclidean or "distance" metric). A set is convex if the straight line connecting any two points in the set is completely contained in the set. This condition excludes sets with jagged boundaries, or sets with holes. A major advantage of our approach over the older complex combinatorial geometry (CCG) packages is that the component objects of a region are allowed to share boundaries.

While these mathematical requirements on region-hood are restrictive, they basically reflect our intuitive concept of fabricated material body with smooth surfaces. All such geometric systems encountered in the study of radiation dose calculations easily satisfy these requirements.

The three basic functional features of a geometric modeling system are strictly separated in our implementation of volume-based solid modeling, in contrast to the CCG systems of the public domain codes. To perform ray tracing for a given trajectory, our algorithm proceeds as follows. First, the code interrogates each region needed to describe the subject geometry, starting with the first or outermost region, and checks for intersections of the ray with its boundary. If intersections are found, the algorithm then interrogates the list of regions $\mathrm{INT}(R_1)$. The scope of the search is restricted by interrogating only those lists of regions $\mathrm{INT}(R_i)$ for which the trajectory intersects the boundaries of R_i. Interrogation of a region R_i consists of calculating the intersections of the ray with the boundaries of the n_i component objects $\{O_{i1},...,O_{ij},...,O_{in_i}\}$ comprising R_i. For example, if a particular component is an ellipsoid centered at the origin with X-, Y-, and Z-axis radii a, b, and c, respectively, intersection finding consists of solving the equation

$$\left(\frac{x_0+\Omega_x\cdot d_k}{a}\right)^2+\left(\frac{y_0+\Omega_y\cdot d_k}{b}\right)^2+\left(\frac{z_0+\Omega_z\cdot d_k}{d_k c}\right)^2=1 \quad (1)$$

for d_k, where the origin of the trajectory is denoted by (x_0,y_0,z_0) and its direction is given by its direction cosines $(\Omega_x,\Omega_y,\Omega_z)$. In practice, such root finding is accelerated by applying constraints to these equations, e.g., that the distance to intersection be real, positive, and less than the maximum length of the flight path. The result of interrogating the component objects of a region is a list of intersections. Let this list be denoted by $\{\mathbf{e}_1,\mathbf{e}_2,\mathbf{e}_3,...,\mathbf{e}_{m_i}\}$, where m_i is the total number of intersections of the trajectory with the surfaces O_{ij} that define the boundary of region R_i, with each $\mathbf{e}_k$ corresponding to a solution d_k of equations of form as Eq. (1). Only a subset of this list represents valid intersections with the boundary of the complex region R_i. If the component objects O_{ij} share any common boundaries, the same distance to intersection d may appear multiple times in the list. The next step is to identify the subset consistent with the sequence of set theoretic operators $\{C_{i1},...,C_{ij},...,C_{i_{ni}}\}$.

The result of the above search for intersections is an ordered list (in order of ascending length d_k) of intersections $\{\mathbf{e}_1,\mathbf{e}_2,\mathbf{e}_3,...,\mathbf{e}_m\}$, where m is the total number of intersections of the trajectory with any region boundaries and d_k is the distance from the trajectory origin $\mathbf{r}_0$ to $\mathbf{e}_k$. The final step is to perform topological segment assignment: Given a list of ray-path segments $\{l_1, l_2, l_3,..., l_{m-1}\}$, where $l_k = d_{k+1} - d_k$, produce a list of regions $\{R_{B_1}, R_{B_2}, R_{B_3},..., R_{B_{m-1}}\}$ containing each of these disjoint segments. This is achieved by appropriately processing the list $\{\mathrm{EXT}(R_{B_1}), \mathrm{EXT}(R_{B_2}), \mathrm{EXT}(R_{B3}),..., \mathrm{EXT}(R_{B_{m-1}})\}$, which defines topological location of the segment end points within the system geometry in terms of the relation "is immediately contained in." This processing algorithm must be capable of treating nonconvex regions and common boundaries between regions.

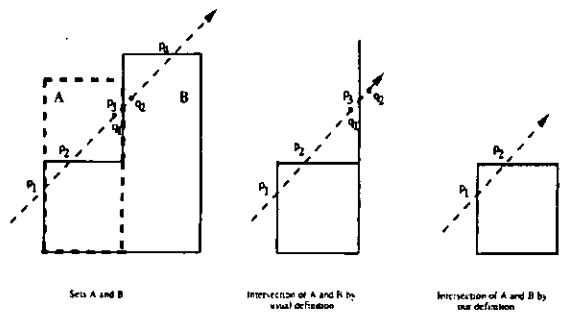


FIG. 2. Dangling boundary resulting from applying the traditional definition of set boundary to intersection of sets A and B; Application of the neighborhood test removes the dangling boundary.

We now consider in detail the problems of (a) processing intersections with boundaries of component objects to identify intersections with boundaries of set-theoretically defined complex bodies and (b) segment classification in a nested structure of nonconvex complex regions.

Combinatorial geometric modeling is available in several public-domain radiation-transport code packages. Two approaches exist. MCNP[4] models geometric objects as intersections, unions, and differences of sets defined by three-dimensional surfaces. This approach provides great versatility in defining arbitrarily complex geometric bodies. It can however be difficult to use. For example, to model a simple rectangular prism one must specify equations of all the six planes that define its surfaces. The TIGER[5] and MORSE-CG[7] code packages, on the other hand, model geometric bodies as topological combinations of primitive geometric objects. However both packages lack the ability to handle objects with coinciding boundaries. By developing proper boundary-point test algorithms and set intersection, union, and difference algorithms, we are able to create a geometric modeling code package that is easy to use and handles the coinciding boundary problem.

B. Intersection identification for set-theoretically defined regions

Figure 2 illustrates the problem of identifying the intersection of a ray with the boundary of region $A \cap B$ from the intersections with the boundaries of its component bodies A and B. If A and B have common boundary, and if the boundary of the complex region is defined to be the usual

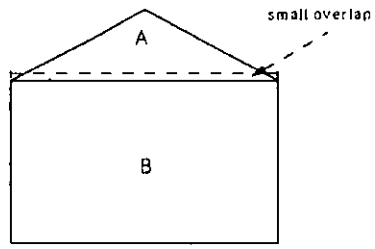


FIG. 3. A case where the approach adopted by MAGI public domain codes fail to solve the dangling boundary problem. The overlap added to either set A or B leads to erroneous description of the geometric structure formed as the union of sets A and B.

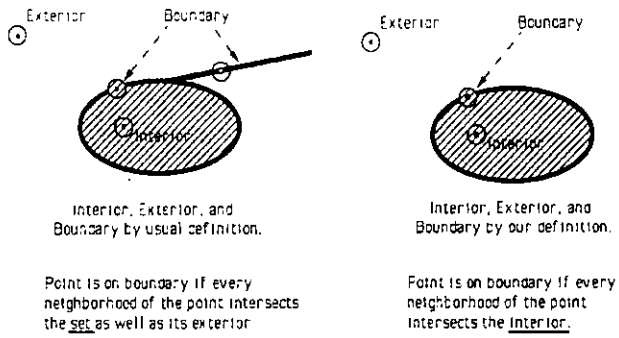


FIG. 4. Revised definitions of set boundary, interior, and exterior. These definitions precludes the existence of dangling boundaries.

topological intersection of the component-object boundaries, i.e., $\partial(A\cap B) = [\partial(A)\cap\partial(B)]\cup[\partial(A)\cap B] \cup[\partial(B)\cap A]$, the so-called "dangling boundary" problem is encountered. As shown in Fig. 2, dangling boundaries are sub-3-D surfaces or lines consisting of boundaries that have collapsed onto themselves and do not surround any interior points. Such pathological intersections must be eliminated, since they confuse segment-classification algorithms, which rely on the fact that segments connecting adjacent intersections with a complex region strictly alternate between being inside and outside the region. The public-domain CCG packages solve this problem by requiring that no such common boundaries exist. This is achieved by using small overlaps such that each overlap extends into one and only one component.[8] Not only does this greatly increase the complexity of geometry specification, but it is in some cases impossible. As shown in Fig. 3, if the top boundary of B is extended into A, the overlap of B will be partly inside and partly outside A. To solve this problem, we have developed a set-theoretically unconventional but rigorously consistent definition of boundary that automatically excludes dangling boundaries under set-theoretic combination of objects with common boundaries.

In the language of set theory, we define the interior, exterior, and boundary points of a set S, constructed from a finite number of convex sets with nonempty interiors, as follows.

The interior of a set S is given by $I(S) = \{p\in E^3 \mid \exists\epsilon>0, S_\epsilon(p)\subset S\}$,

The boundary of S is given by $\partial(S) = \{p\in E^3 \mid p\notin I(S)$ and $\forall\epsilon>0,\ S_\epsilon(p)\cap I(S)\neq\emptyset\}$.

The exterior of S is given by $E(S) = \{p\in E^3 \mid p\notin I(S)$ and $p\notin\partial(S)\}$, E^3 denotes 3-D Euclidean space, and $S_\epsilon(p)$ denotes the open neighborhood or ball of radius ϵ about p: $S_\epsilon(p) = \{p' \mid p'\in E^3$ and $|p-p'|<\epsilon\}$.

In words, these definitions say that the interior of an object S is the largest open set contained in S and that its exterior is the set of points disjoint from the interior and boundary of S. A point p is on the boundary of S that contains nonempty interior if and only if every open ball centered at p overlaps both the interior and exterior of S. In contrast, the conventional topological definition says that every open ball about a boundary point must overlap both the exterior of S (as defined conventionally) and S itself (which can include boundary points). It is obvious that our revised definition, but not the conventional definition, excludes dangling or sub 3-D edges from $\delta(S)$, for the type of sets S under consideration. These definitions are illustrated in Fig. 4.

Next, we derive some basic properties of set intersection, union, and difference based on the above definitions.

Lemma 1: $I(A\cap B) = I(A)\cap I(B)$.

[*Proof:* This is obvious since a small open ball in $A\cap B$ must be contained in A and B, respectively, and vice versa.]

Theorem 1: For $C = A\cap B$, $\partial(C) = \{p\in\partial(A)\cup\partial(B) \mid \forall\sigma>0, S_\sigma(p)\cap I(A)\cap I(B)\neq\emptyset\}$.

[*Proof:* (1) If $p\in\partial(A)\cup\partial(B)$, then $p\notin I(C) = I(A)\cap I(B)$ by Lemma 1. If in addition $\forall\sigma>0$, $S_\sigma(p)\cap I(A)\cap I(B)\neq\emptyset$, then $p\in\partial(C)$ by definition of $\partial(C)$.

(2) We now show that if $p\in\partial(C)$, then $p\in\partial(A)\cup\partial(B)$ and $\forall\sigma>0$, $S_\sigma(p)\cap I(A)\cap I(B)\neq\emptyset$. The universal set E^3 can be easily proven to be the union of four disjoint sets:

$$E^3 = [I(A)\cap\sim I(C)]\cup[I(B)\cap\sim I(C)]$$

$$\cup[\sim I(A)\cup\sim I(B)]\cup I(C),$$

where $\sim I(A)$ is the complement of $I(A)$ in E^3. Any point p on the boundary $\partial(C)$ of set C must fall into one of the first three sets since $p\in\partial(C)$ implies that $p\notin I(C) = I(A)\cap I(B)$ by Lemma 1. Examining the other possibilities, we have (a) If $p\in I(A)\cap\sim I(C)$, then $p\notin I(B)$ and $p\notin\partial(A)$. Since $p\in\partial(C)$ implies that $\forall\sigma>0$, $S_\sigma(p)\cap I(B)\neq\emptyset$, then $p\in\partial(B)\cap\sim\partial(A)$. (b) If $p\in I(B)\cap\sim I(C)$, then $p\in\partial(A)\cap\sim\partial(B)$ by similar reasoning. (c) If $p\in\sim I(A)\cup\sim I(B)$, then $p\in\partial(A)\cap\partial(B)$.

Since $\partial(A)\cup\partial(B) = [\partial(B)\cap\sim\partial(A)]\cup[\partial(A)\cap\sim\partial(B)]\cup[\partial(A)\cap\partial(B)]$, it follows that $\partial(C)\subset\partial(A)\cup\partial(B)$. Q.E.D.]

Theorem 1 shows that the boundary $\partial(C) = \partial(A\cap B)$ can be decomposed into three disjoint components:

$$\partial(C) = \{[\partial(B)\cap\sim\partial(A)]\cap[I(A)\cap\sim I(C)]\}$$
$$\cup\{[\partial(A)\cap\sim\partial(B)]\cap[I(B)\cap\sim I(C)]\}$$
$$\cup\{p\in\partial(A)\cap\partial(B)\,|\,\forall\epsilon>0,$$
$$S_\epsilon(p)\cap I(C)\neq\emptyset\}.$$

This finding yields the following algorithm for identifying the boundary of the intersection of A and B:

(1) If p_i is in $\partial(A)$ or $\partial(B)$, but not in both, then $p_i\in\partial(A\cap B)$ if (and only if) p_i is in $I(B)$ or $I(A)$, respectively;

(2) Else if $p_i\in\partial(A)\cap\partial(B)$, then perform the neighborhood test. Choose q_1 and q_2 in a small neighborhood of p_i:

If $q_1\in I(A\cap B)$ and $q_2\notin I(A\cap B)$ or

$q_2\in I(A\cap B)$ and $q_1\notin I(A\cap B)$

then

$p_i\in\partial(A\cap B)$,

else

$p_i\notin\partial(A\cap B)$,

endif.

Note that the above algorithm is only an approximate implementation of Theorem 1. In our implementation, q_1 and q_2 are points along the ray path located at small distance δ proximal and distal to the intersection being evaluated. It nevertheless effectively removes the "dangling boundary" as illustrated by Fig. 2. Experience shows that setting δ to 10^{-3} mm yields good results when VAX single-precision floating-point arithmetic (approximately seven significant digits) is used. In general, comparisons of floating-point numbers must be handled carefully in order to avoid round-off and truncation errors. Comparisons of the form (A .eq. B) must be replaced by control statements of the form ($|A-B|$ .le. ϵ) where ϵ is a small number (1.5×10^{-6} in our code) that represents "machine zero." In addition, implementation of our intersection and union algorithms requires answering questions of the form "does point p, which denotes an intersection with $\partial(A)$, lie on the boundary of some other object B?" Finite register-length problems are avoided by assuming that the boundary of each region is a thin "shell" of thickness 10ϵ.

Similar algorithms have been developed for the implementation of set unions and differences. We omit the discussion of these algorithms here for the sake of brevity. The following algorithm finds the boundary of the union of sets A and B:

(1) If p_i is in $\partial(A)$ or $\partial(B)$, but not in both, then $p_i\in\partial(A\cup B)$ if (and only if) p_i is not in $I(B)$ or $I(A)$;

(2) Else $p\in\partial(A)\cap\partial(B)$, and perform the neighborhood test. Choose q_1 and q_2 in a small neighborhood of p_i:

If $q_1\in I(A\cup B)$ and $q_2\notin I(A\cup B)$ or

$q_2\in I(A\cup B)$ and $q_1\notin I(A\cup B)$,

then

$p_i\in\partial(A\cup B)$,

else

$p_i\notin\partial(A\cup\beta)$,

endif.

It should be noted that our algorithms achieve the same effect as the regularized set theory by Requicha and Tilove.[9,10]

C. Segment classification

Having obtained the intersection points of a given ray with the boundary of a geometric region defined as set intersections, unions, or differences of primitive geometric objects, we now consider the problem of identifying the region containing the segment bounded by two adjacent intersections. A finite set of regions $\{R_j\}$, $j=0,\ldots,n$, is used to represent any complex physical system. Each region R_j is assumed to have orientable surfaces and to be set-theoretically defined from finite numbers of convex geometric objects. We assume that the outermost region, which defines the spatial "universe" of the problem, is the set R_0. The sets $\{R_j\}$ have the following properties:

(1) $R_i\subset R_0$, $i=0,\ldots,n$;

(2) if $j\neq i$, then either $I(R_i)\cap I(R_j)=\emptyset$, or $R_i\subset R_j$. That is, the interiors of sets R_i are either disjoint or the sets are wholly contained in one another.

Intersections of a ray with the boundaries of regions R_i can be represented as a list $\{\mathbf{e}_j\}=\{(\mathbf{r}_1+d_j\mathbf{\Omega}_1)\}$ ordered in terms of ascending d_j. Intersections of the ray path with coinciding boundaries of multiple regions are listed once for each boundary. The order of identical shared-boundary intersections is determined by the INT(R_i) arrays, the order of region definition in the input file and the direction of the ray. For any two adjacent intersection points $\mathbf{e}_j$ and $\mathbf{e}_{j+1}$ let B_j and B_{j+1} denote the number of the region intersected, i.e., $\mathbf{e}_j\in\partial(R_{B_j})$ and $\mathbf{e}_{j+1}\in\partial(R_{B_{j+1}})$. Letting $l=B_j$ and $m=B_{j+1}$, the relations between $I(R_l)$ and $I(R_m)$ can be enumerated as $I(R_l)\cap I(R_m)=\emptyset$, $I(R_m)\subset I(R_l)$ or $I(R_l)\subset I(R_m)$.

A region R as is considered in this context can be represented as a surface in the three-dimensional Euclidean space, denoted by a continuous function F such that $F(x, y, z)=0$. The requirement that region boundaries be orientable implies the existence of function $F(x, y, z)$ such that $F(x, y, z)<0$ on one side of the boundary and $F(x, y, z)>0$ on its other side. We present a ray as a path in the three-dimensional Euclidean space, $[a(t), b(t), c(t)]$, $0<t<T$, for some $T>0$. Then the intersections of the path with the boundary $\partial(A)$ are the solutions of the equation $C(t)=0$:

$$C(t)\equiv F[a(t),\ b(t),\ c(t)]=0,\ 0<t<T.$$

Since F, a, b, and c are all continuous functions, the function $C(t)$ is continuous in t. Note that any point $[a(t_0), b(t_0), c(t_0)]$ is outside of the surface if $C(t_0)>0$ and is inside the surface if $C(t_0)<0$.

Using the above notations, we define an intersection as follows,

An intersection of the path $[a(t), b(t), c(t)]$, $0 \leqslant t \leqslant T$, with the surface $F(x, y, z) = 0$ is a root t_0 of the function $C(t)$, $0 \leqslant t \leqslant T$, if there exists an open interval around t_0 such that $C(t) > 0$ for $t > t_0$ and $C(t) < 0$ for $t < t_0$ (or the other way around) when $t_0 \neq 0$ and $t_0 \neq T$. In the cases of $t_0 = 0$ or $t_0 = T$, $C(t)$ is less than (or greater than) 0 in a small neighborhood near t_0 in the range
of t.

By the above definition of intersection, we have excluded "grazing" intersections of the path with the surface of region R, as well as the nontrivial coincidence of the path with the surface of region R, i.e., the existence of points t_1 and t_2, $t_1 \neq t_2$, such that $C(t) = 0$ for all t between t_1 and t_2.

To derive the algorithm for classifying the segment between intersections $\mathbf{e}_i$ and $\mathbf{e}_{i+1}$, we need the following lemmas.

Lemma 2: Let $\mathrm{EXT}(R_i) = \{\cap R_j | i \neq j$ such that $R_i \subset R_j\}$, then $\mathrm{EXT}(R_i)$ is the smallest nonempty set in $\{R_j\}$ that immediately contains R_i.

[*Proof:* (1) Since $R_i \subset R_0$, the set $\mathrm{EXT}(R_i)$ is nonempty.

(2) Since any two sets R_s and R_t containing R_i satisfy the condition $R_s \subset R_t$ or $R_s \subset R_p$ there exists a unique ordering of such sets R_1, R_2,..., R_k such that $R_1 \subset R_2 \subset \cdots \subset R_k$. Therefore, $\mathrm{EXT}(R_i) \in \{R_j\}$, and contains R_i immediately since it is the smallest set containing R_i.]

Lemma 3: Assume that the distance between neighboring intersections $\mathbf{e}_i$ and $\mathbf{e}_{i+1}$ is nonzero. If $R_l \subset R_m$, $\mathbf{e}_i \in \partial(R_l)$, and $\mathbf{e}_{i+1} \in (R_m)$, then $R_m = \mathrm{EXT}(R_l)$.

[*Proof:* If the lemma is not true, then there exists R_k such that $R_l \subset R_k \subset R_m$, $R_l \neq R_k$, and $R_k \neq R_m$. Thus the line segment from $\mathbf{e}_i$ to $\mathbf{e}_{i+1}$ intersects the boundary of R_k, which is a contradiction since $\mathbf{e}_i$ and $\mathbf{e}_{i+1}$ are adjacent intersection points with nonzero distance between them.]

We are now ready to obtain our segment classification algorithm.

Theorem 2: Two different cases can be found for the segment classification problem. In the first one, neighboring intersection points are located on the boundary of the same region, say R_m. This is handled by part (1) of the theorem. In the second case, neighboring intersection points belong to boundaries of different regions and are handled by part (2) of the theorem.

(1) Let $l = B_i$, $m = B_{i+1}$. Assume $l = m$, i.e., the intersections $\mathbf{e}_i$ and $\mathbf{e}_{i+1}$ are both on the boundary of region m. Thus we restrict our attention to region R_m only. Assume that the ray starts from outside of R_m. We construct from the set of intersections of the ray $\{\mathbf{e}_i\}$ with boundaries of all regions a subset of intersections $\{\mathbf{g}_j\}$, $j = 1,..., n_m$, containing n_m intersections with the boundary of region R_m only. Then the line segment $(\mathbf{g}_{2j}, \mathbf{g}_{2j+1}) \subset \mathrm{EXT}(R_m)$ and $(\mathbf{g}_{2j+1}, \mathbf{g}_{2j+2}) \subset R_m$. The oddity of $\mathbf{g}_j$ is reversed if the ray starts from inside of the region R_m.

(2) If $l \neq m$, $\mathbf{e}_i \in \partial(R_l)$, $\mathbf{e}_{i+1} \in \partial(R_m)$, and distance between $\mathbf{e}_i$ and $\mathbf{e}_{i+1}$ is nonzero, then (a) If $I(R_l) \cap I(R_m) = \varnothing$, then $(\mathbf{e}_i, \mathbf{e}_{i+1}) \subset \mathrm{EXT}(R_m \cup R_l) = \mathrm{EXT}(R_m) = \mathrm{EXT}(R_l)$. (b) If $R_l \subset R_m$, then $(\mathbf{e}_i, \mathbf{e}_{i+1}) \subset R_m$. (c) If $R_m \subset R_l$, then $(\mathbf{e}_i, \mathbf{e}_{i+1}) \subset R_l$.

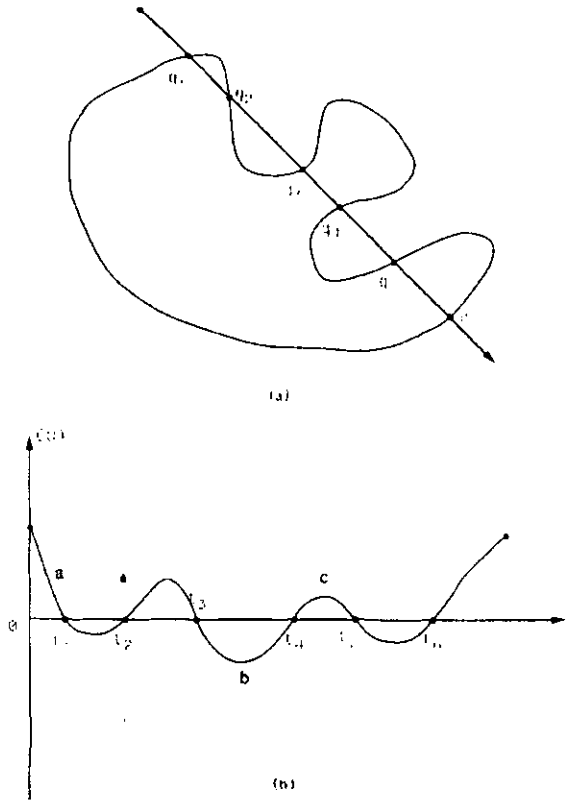


FIG. 5. Proof of our segment classification algorithm. The ray path and the region boundary (a) are transformed to a function $C(t)$, (b) whose roots are intersections of the ray path with the region boundary.

[*Proof:* (1) Representing the ray as a function $[a(t), b(t), c(t)]$, $0 \leqslant t \leqslant T$, and the region R_m by a function $F(x, y, z) = 0$, we see that the intersections $\mathbf{g}_j$'s correspond to the roots of the function $C(t)$ as was defined before. Let $\{t_j\}$, $j = 1,..., n_m$, be the set of roots of the function $C(t)$, ordered according to their distance to 0, and set $t_0 = 0$. This is possible due to the fact that the region enclosed by $F(x, y, z)$ is made up of a finite number of convex sets. Since the ray starts from outside of the region R_m, we have $C(t_0) > 0$. We show by mathematical induction that (a) $C(t) > 0$ for $t_{2i} < t < t_{2i+1}$, and (b) $C(t) < 0$ for $t_{2i+1} < t < t_{2i+2}$. As shown in Fig. 5, this transforms the original segment classification problem in 3-D space into a 1-D problem.

Any segment of the function $C(t)$ between its roots, say t_i and t_{i+1}, can be of only one sign, either positive or negative, i.e., the segment is either completely inside or outside the region. If this were not true, then there would exist at least two points s_1 and s_2, $t_i < s_1 < s_2 < t_{i+1}$, such that $C(s_1) > 0$ and $C(s_2) < 0$ or $C(s_1) < 0$ and $C(s_2) > 0$. Since $C(t)$ is continuous, then there must exist s_0 such that $s_1 < s_0 < s_2$ and $C(s_0) = 0$ by the extended mean value theorem.[11] This contradicts our hypothesis that t_i and t_{i+1} are neighboring intersections with no other intersection between them.

Now we look at the segment of the curve $C(t)$ denoted

by a in Fig. 5 (b). Since $C(t_0) > 0$, this segment is outside of the region by the above conclusion. Since, by our definition of an intersection, there must exist points s_1 and s_2 near t_1 such that $C(s_1) > 0$ and $C(s_2) < 0$, it follows that s_2 must lie between t_1 and t_2. Therefore, segment denoted by b must lie inside the region. Similarly we can prove that the segment denoted by c lies outside of the region. It is trivial to apply mathematical induction to arrive at the conclusions of (a) and (b) above.

(2a) We first show that $\mathrm{EXT}(R_m \cup R_l) = \mathrm{EXT}(R_m) = \mathrm{EXT}(R_l)$. By definition, $\mathrm{EXT}(R_m) = \{\cap R_i | i \neq m$ such that $R_m \subset R_i\}$, and $\mathrm{EXT}(R_m \cup R_l) = \{\cap R_i | i \neq m, i \neq l$, such that $R_m \cup R_l \subset R_i\}$. It is obvious that $\mathrm{EXT}(R_m) \subset \mathrm{EXT}(R_m \cup R_l)$. Since $I(R_l) \cap I(R_m) = \emptyset$, $\mathrm{EXT}(R_m \cup R_l) \neq \mathrm{EXT}(R_m)$ if and only if there exists an R_j in the sets defining $\mathrm{EXT}(R_m)$ such that R_j does not contain R_l. However, if this were true, then the line segment from $\mathbf{e}_i$ to $\mathbf{e}_{i+1}$ would intersect the boundary of R_j, which is a contradiction since $\mathbf{e}_i$ and $\mathbf{e}_{i+1}$ are neighboring intersection points with nonzero distance between them. Therefore, we have $\mathrm{EXT}(R_m \cup R_l) = \mathrm{EXT}(R_m)$. Similarly, $\mathrm{EXT}(R_m \cup R_l) = \mathrm{EXT}(R_l)$.

(2b) Since $R_l \subset R_m$, $\mathrm{EXT}(R_l) \subset R_m$. Since $\mathbf{e}_i \in \partial(R_l)$, and $\mathbf{e}_{i+1} \in \partial(R_m)$ are neighboring intersections with nonzero distance between them, there cannot exist another set R_j such that $R_l \subset R_j \subset R_m$, therefore R_m is the smallest set immediately containing R_l.

(2c) This is simply the reversal of (2b) hence we omit the proof. Q.E.D.]

Based on the above proposition, the algorithm for segment classification, assuming no coinciding region boundaries, can be formulated as the following table of rules (see Fig. 6):

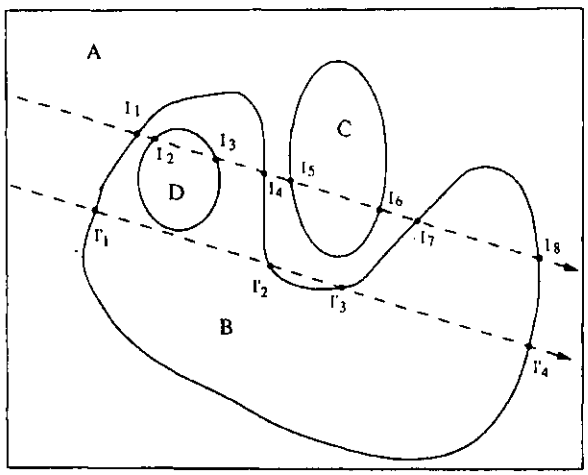


FIG. 6. An example of the application of our segment classification algorithm.

Let $\mathbf{e}_i = i$th intersection of the ray with the region boundaries ordered with increasing distance from source. Assume that $\mathbf{e}_i$ is located on R_{B_i}, $n_{B_{i-1}}$ is the number of intersection points of the ray with boundary of $R_{B_{i-1}}$ among $\{\mathbf{e}_j\}, j = 1, i-1$:

Case	Assignment	Example (Fig.6)
$R_{B_i} = R_{B_{i-1}}$, $n_{B_{i-1}}$ odd	segment $(\mathbf{e}_{i-1}, \mathbf{e}_i)$ in R_{B_i}	(I_2,I_3), (I_5,I_6), (I_7,I_8), (I'_1,I'_2), (I'_3,I'_4)
$R_{B_i} = R_{B_{i-1}}$, $n_{B_{i-1}}$ even	segment $(\mathbf{e}_{i-1}, \mathbf{e}_i)$ in $\mathrm{EXT}(R_{B_i})$	(I'_2,I'_3)
$\mathrm{EXT}(R_{B_{i-1}}) = \mathrm{EXT}(R_{B_i})$	segment $(\mathbf{e}_{i-1}, \mathbf{e}_i)$ in $\mathrm{EXT}(R_{B_i})$	(I_4,I_5), (I_6,I_7)
$R_{B_{i-1}} = \mathrm{EXT}(R_{B_i})$	segment $(\mathbf{e}_{i-1}, \mathbf{e}_i)$ in $R_{B_{i-1}}$	(I_1,I_2)
$\mathrm{EXT}(R_{B_{i-1}}) = R_{B_i}$	segment $(\mathbf{e}_{i-1}, \mathbf{e}_i)$ in R_{B_i}	(I_3,I_4)

The actual implementation of segment classification requires treatment of the case when the ray originates from a region other than the universal region E^3. In addition, the case when regions have "touching" boundaries is much more complex to handle. This is handled in our implementation by several heuristic criteria.

A critical difference between the above algorithm and the Siddon algorithm[12] is that the latter only finds the total radiological distance along a ray in a given system, while our algorithm produces an array of ray segments sequentially ordered according to distance from the origin of the ray and an associated array of medium numbers identifying the medium containing each of the ray segments. Many transport calculations require this additional information, including the Monte Carlo problem of randomly choosing a collision point along a fixed trajectory.

The "topography algorithm" by Bistrovic *et al.*[13] performs the same function as our segment classification algorithm. In the topography algorithm, a stack is created to keep track of all intersection points with the geometric regions. These intersection points are "removed" from the stack by pairs according to the regions that are intersected.

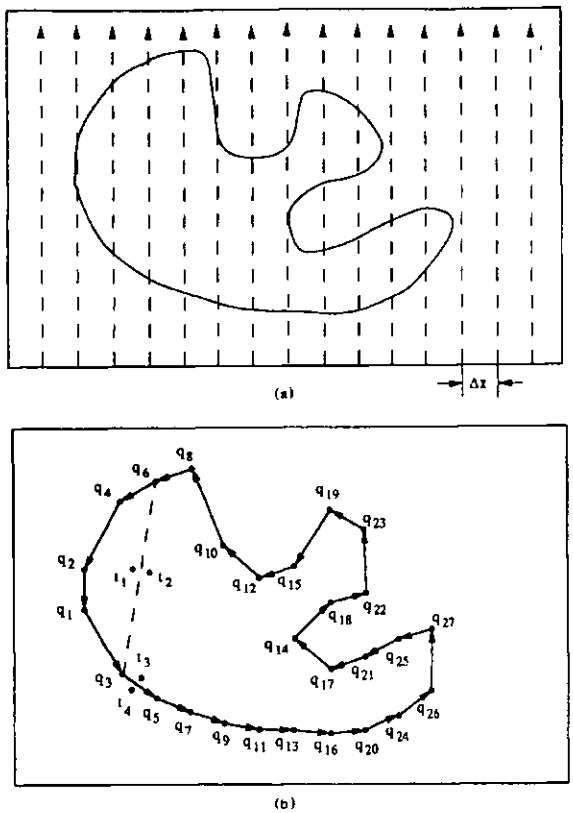


FIG. 7. An example of the application of our iterative local boundary point search (chain code) algorithm. (a) shows how the 2-D unordered grid of points, $\{q_i\}$, is generated by tracking a set of parallel rays contained in image plane. (b) Illustrates how the chain-code algorithm reorders these points so that an incremental pattern can be used to correctly draw the boundary.

No mathematical proof is presented, although an argument similar to the one in the proof of (1) in Theorem 2 can be used.

D. Graphic verification of geometric models

The complexity of geometric models that can be constructed with our geometric modeling code package often challenges human intuition. Graphic tools that visualize complex geometric models have proven to be very helpful in verifying their accuracy and consistency. We have developed two such tools for this purpose. One finds the intersection of a geometric model with a user-specified plane in the 3-D space, and draws the outline of this intersection on a graphic display device (currently a Silicon Graphics IRIS graphics workstation). The other tool displays a simulated radiograph of the model in three orthogonal views. We discuss these tools in the following.

We first discuss the algorithm to find the planar intersection of a given geometric model. Given a plane specified by its normal vector (n_1, n_2, n_3), the intersection points $\{q_i\}$, $i = 1,..., n$, of the boundary of a particular region in the geometric model with the plane are computed, as shown in Fig. 7(a), by finding the intersections of a series of trajectories in the plane all parallel to the y axis of the plane, such that neighboring rays are separated by Δx. In the display-plane coordinate system, we denote the coordinates of the q_i's by (x_i, y_i) where $x_{i+1} = x_i + \Delta x$. The problem then becomes one of reordering the set $\{(x_i, y_i)\}$, or "connecting the dots," such that each pair of adjacent boundary points is connected by a line segment so that the boundary can be displayed by drawing a sequence of 2-D bounded vectors. To our knowledge, no chain code algorithm which exploits the topological properties of "inside/outside" exists in the published literature. This is done, for a given current boundary point, by selecting a candidate adjacent point on the same boundary and performing a check to ensure that it is indeed the next boundary point according to our definition of boundary. Since all region boundaries are continuous, their intersections with the arbitrary plane are piece-wise continuous.

Let a piece of the intersection of region R boundary with the arbitrary plane be denoted by a curve G, then $G = G(t)$, for t in the interval $[t_0, T]$, and $G(t)$ is continuous. Given a point $p_1 = G(t_1)$ on G, its adjacent point is defined as point p_2 such that $p_2 = G(t_2)$, $t_2 > t_1$ and there does not exist $q_j = G(t_j)$ in $\{q_i\}$, $i = 1,..., n$, such that $t_1 < t_j < t_2$.

The determination of an adjacent point is performed by a "neighborhood test." Given a current point p_i, a set of candidate adjacent points are chosen such that their x-axis coordinate values are one Δx away from, or the same as, that of p_i. Given a candidate adjacent point s of current point p_i selected from such set, we selected two test points on the perpendicular bisector of the line connecting points p_i and s [for example, points t_1 and t_2 in Fig. 7(b)], and identify the regions that these test points belong to. If the test points belong to the same region, then the candidate point cannot be the adjacent point, and it is rejected. Otherwise the candidate point is saved in a list, while more candidate adjacent points are tested. Finally, the candidate adjacent point with the smallest distance to p_i is taken to be p_{i+1}.

Heuristic selection of the test points t_1 and t_2 is implemented. The distance ϵ from the test points to the line linking points p_i and s, is specified by the user for best results. The two triangles formed by p_i, t_1, t_2, and s, t_1, t_2 are identical.

Let $\{q_i\}$ denote the unordered list of boundary points and $\{p_i\}$ the list of boundary points that have been ordered so far. We have developed an iterative local search algorithm based on this continuity property as follows:

(1) If $\{p_i\}$ is empty, choose $p_0 = (x_0, y_0) = q_0$, such p_0 is an intersection point with the smallest x-axis coordinate value. Let the piece of region boundary that contains p_0 be denoted by $G = G(t)$, $t_0 \leqslant t \leqslant T$, indexed by l.

(2) Assuming that p_i is on the boundary of R, $p_1,..., p_i$ are ordered so that $p_0 = G(t_0),..., p_i = G(t_i)$, $t_0 < t_1 < ... < t_i$, and p_{i+1} is the adjacent point of p_i on G. Then

(a) Test all unordered points in the set $\{g \in \{q_i\} | g = (x_i + \Delta x, y)\}$, where x_i is the x-coordinate of p_i. The g's are candidate adjacent points. If multiple points in this set are found to satisfy the neighborhood test, the point with the smallest distance to p_i is taken as the adjacent point. If an adjacent point p_{i+1} of p_i is found, increase

search index and return to (a), else set flat = LEFT and go to (b);

(b) Test all unordered points in the set $\{g \in \{q_i\} \mid g = (x_i, y)\}$.

If p_{i+1} is found, then increase search index and:
  if flag = LEFT,
    go to (a),
  else
    go to (c).
  endif
Else
  if flag = LEFT,
    go to (c),
  else
    go to (a),
  endif
Endif.

(c) Test all unordered points in the set $\{g \in \{q_i\} \mid g = (x_i - \Delta x, y)\}$. If an adjacent point p_{i+1} is found, increase search index and return to (c), else set flag = RIGHT and go to (b).

To prevent an endless loop within this step, a stack is created to store the last 5 values of the flag. The stack is checked at the beginning of substeps, a, b, and c. If it is found that a complete loop has been performed among the three substeps without finding the adjacent point, the program leaves this step and goes to step 3.

The flags LEFT and RIGHT identifies the location of the current point relative to the candidate adjacent point, i.e., on its left or right.

(3) If the search of (2) fails, the lth continuous piece of the boundaries has been found. Increment l by 1, go to next available q_j and start searching for another continuous piece of the boundary. The program exits if no more points are to be ordered.

A flow chart of the above algorithm is shown in Fig. 8.

As an example, Fig. 7(b) shows how the algorithm works for the region boundary in Fig. 7(a). Starting from point q_1, the algorithm proceeds to test points q_3 and q_4, which are immediately to the right of q_1, according to part (a) of the algorithm, and finds q_3 to be the next point on the boundary. The algorithm proceeds this way until it finds point q_{26} to be the next boundary point from q_{24}. Here part (b) of the algorithm is executed and point q_{27} is found to the next boundary point. The algorithm then switches to execute part (c) of the algorithm to find boundary points q_{25}, q_{21}, q_{17}, q_{14}, in this order, and finds q_{12} to not be the next boundary point from q_{14}. Part (b) of the algorithm is executed again at this point to test points q_{15} and finds it not to be the next boundary point since this fails the neighborhood test. Therefore part (a) of the algorithm is executed again to test points q_{18} and q_{19}. This process continues until all points q_1 through q_{27} are tested and ordered.

The above local search algorithm, compared with a global search algorithm, results in tremendous savings of cpu time. A global search algorithm, which tests every remaining intersection points to find the next point on the boundary segment, would always require $(n-1)!$ tests to completely search through the remaining n unordered intersection points. Our local search algorithm, in most common cases where a simple closed boundary is to be reconstructed, tests less than $3n$ points to completely order all intersection points.

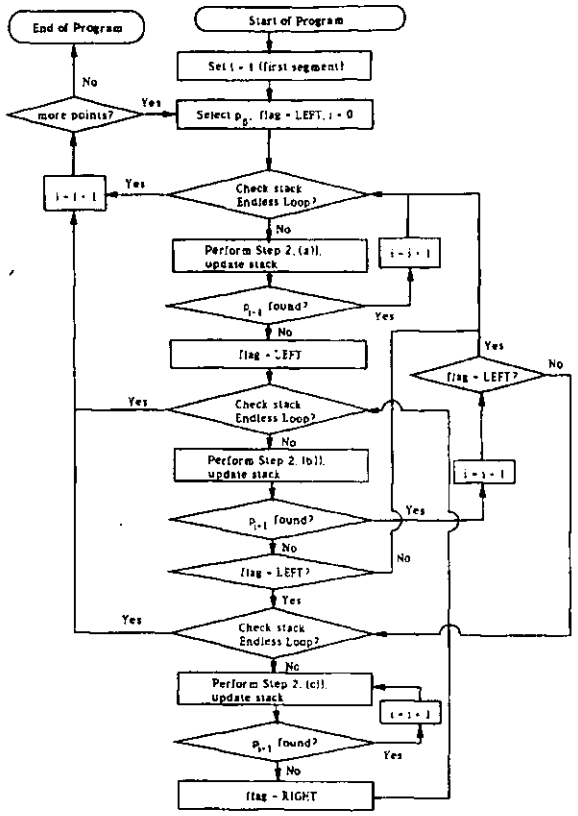


FIG. 8. Flow chart of the "chain code" algorithm.

Results from the above local search algorithm are shown in Figs. 9 and 10. Fig. 9 shows the intersection of a model 6711 I-125 seed[14] with a plane containing its axis. Figure 10 shows the structure of the University of Michigan CT-compatible vaginal colpostat.[15]

Another useful graphic display routine calculates simulated radiographs of a geometric model in three orthogonal views. The cube that contains the model is identified and the image planes perpendicular to the X, Y, and Z axes are divided into pixels. Using the ray-tracing routines described above, the radiological thickness, $\Sigma_{i=1}^{n} \mu_i \cdot d_i$, through the center of each pixel in the direction perpendicular to the image plane is calculated. By calculating the x-ray transmission incident upon each imaging pixel, the optical density yielding a simulated greyscale value. In the computer science literature, this graphic display technique is known as "ray casting." Figure 11 shows the simulated radiographic views of a 3M Fletcher–Suit–Delclos colpostat,[16] while Fig. 12 shows those of the model 6711 I-125 seed. These images are derived from 1-MeV photon attenuation coefficients.

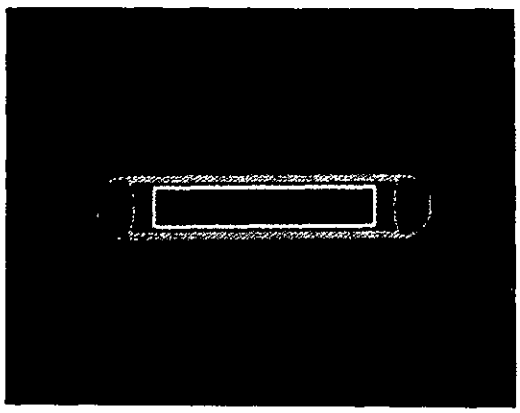


FIG. 9. Intersection of a model 6711 I-125 interstitial seed with a plane containing the seed axis. Elliptical end welds are modeled by representing the titanium capsule as the union of a right cylinder and two spheres and the air cavity as the intersection of a right cylinder and the complements of two spheres.

III. DISCUSSION AND CONCLUSIONS

A set of combinatorial geometric modeling routines written in FORTRAN has been developed based on the mathematical principles shown in the last section. Application of these routines to the modeling of a variety of gynecological applicators, interstitial sources, geometrically complex detectors, and general heterogeneities has been highly successful. Because the underlying intersection, segment classification and point classification algorithms are based upon mathematically rigorous and consistent definitions of boundary and interior, our modeling system is easy to use and has proven to be free of errors and artifacts.

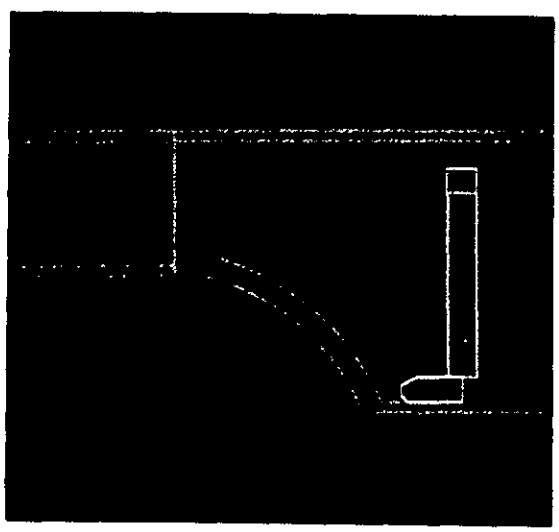


FIG. 10. Intersection of the University of Michigan vaginal colpostat with the plane containing the source and handle axes. The beveled edges of the rectal shield are modeled by representing the shielding as the intersection of the disc with two cones.

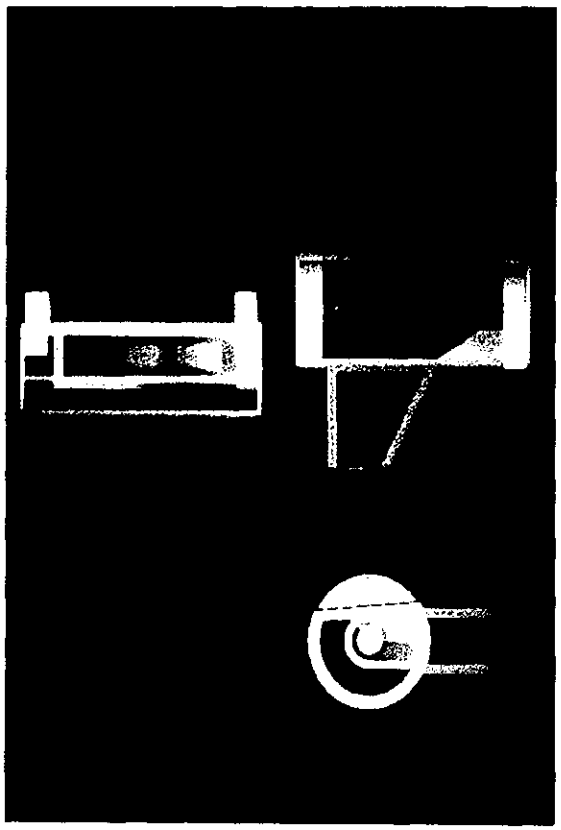


FIG. 11. Simulated orthogonal radiographs of a 3M Fletcher–Suit–Delclos colpostat.

Two auxiliary routines have been developed to help a user verify the geometric model for a given geometry. The first routine computes the projections of a geometric model to three orthogonal planes x-y, x-z, y-z, and displays these projections as grey-scale images on a computer graphics workstation. The second routine displays an arbitrarily oriented cross section of a given geometric model, using ray tracing, and draws the profile in a computer graphics workstation.

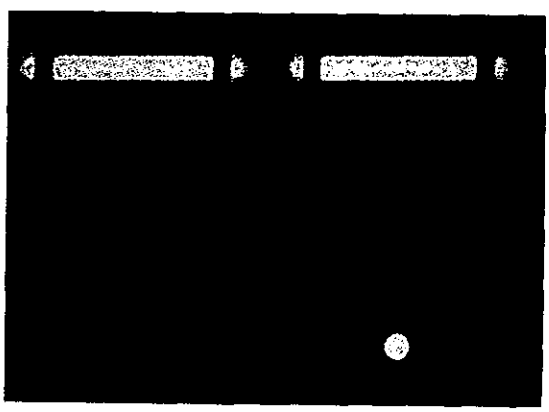


FIG. 12. Simulated radiograph of a model 6711 I-125 interstitial seed.

The generality of our geometric modeling code package cannot be overemphasized. In fact, we have applied it successfully to complement a number of radiation transport calculation algorithms in which geometric models are required, including Monte Carlo photon transport calculations and simple heuristic dose calculation algorithms.

ACKNOWLEDGMENT

This work was supported by NIH/NCI Grant No. R01 CA46640.

a)Address reprint requests and correspondence to: Jeffrey F. Williamson, Ph.D., Radiation Oncology Center, Mallinckrodt Institute of Radiology, Washington University School of Medicine, St. Louis, MO 63110.

[1]J. F. Williamson, "Radiation transport calculations in treatment planning," Comput. Med. Imag. Graph. **13**, 251–268 (1989).

[2]E. E. Lewis, and W. F. Miller, *Computational Methods of Neutron Transport* (Wiley, New York, 1984).

[3]R. L. Siddon, "Fast calculation of the exact radiological path for a three-dimensional array," Med. Phys. **12**, 252–255 (1985).

[4]J. F. Briesmeister, Ed., MCNP—A General Monte Carlo Code for Neutron and Photon Transport, Version 3A., Los Alamos National Laboratory, Los Alamos, New Mexico (1986).

[5]M. O. Cohen *et al.* SAM-CE: A three dimensional Monte Carlo Code for the solution of the forward neutron and forward and adjoint gamma ray transport equations-Revision B, DNA 2830F-B. Mathematical Applications Group, Inc. (1973).

[6]J. A. Halbleib and T. A. Mehlhorn, ITS: The Integrated TIGER Series of Coupled Electron/Photon Monte Carlo Transport Codes. Sandia National Laboratories, Albuquerque, New Mexico (1984).

[7]E. A. Straker, P. N. Stevens, D. C. Irving, and V. R. Cain, "MORSE-CG: General purpose Monte Carlo multigroup neutron and gamma-ray transport code with combinatorial geometry," Radiation Shielding Information Center (ORNL) Report CCC-203 (1976).

[8]T. M. Jenkins, W. R. Nelson, and R. Alessandro, Eds. *Monte Carlo Transport of Electrons and Photons* (Plenum, New York, 1988).

[9]A. A. G. Requicha and R. B. Tilove, "Mathematical Foundations of Constructive Solid Geometry: General Topology of Closed Regular Sets," National Science Foundation Report No. NSF/RA-780069 (1978).

[10]R. B. Tilove, "Set membership classification: A unified approach to geometric intersection problems," IEEE Trans. Comput. **29**(10), 874–883 (1980).

[11]R. C. Buck, *Advanced Calculus* (McGraw-Hill, New York, 1965).

[12]R. L. Siddon, "Calculation of the radiological depth," Med. Phys. **12**, 84–87 (1985).

[13]M. Bistrovic, and T. Viculin, "Use of Computers in Radiation Treatment Planning," Radiologia Iugoslavica **18**(2), 149–154 (1984).

[14]C. C. Ling, E. D. Yorke, I. J. Spiro, D. Kubiatowicz, and D. Bennet, "Physical dosimetry of I-125 seeds of a new design for interstitial implant," Int. J. Radiat. Oncol. Biol. Phys. **9**, 1747–1752 (1983).

[15]S. L. Schoeppel, B. A. Frass, M. P. Hopkins, M. L. LaVigne, A. S. Lichter, D. L. McShan, S. Noffsinger, C. Perez-Tamayo, and J. A. Roberts, "A CT-compatible version of the Fletcher system intracavitary applicator: Clinical application and 3-dimensional treatment planning," Int. J. Radiat. Oncol. Biol. Phys. **17**, 1103–1109 (1989).

[16]J. F. Williamson, "Dose calculations about shielded gynecological colpostats," Int. J. Radiat. Oncol. Biol. Phys. **19**, 167–178 (1990).

INSTITUTE OF PHYSICS PUBLISHING PHYSICS IN MEDICINE AND BIOLOGY

Phys. Med. Biol. **49** (2004) 1991–2013 PII: S0031-9155(04)74385-7

A unifying framework for multi-criteria fluence map optimization models

H Edwin Romeijn[1], James F Dempsey[2] and Jonathan G Li[2]

[1] Department of Industrial and Systems Engineering, University of Florida, Gainesville, Florida 32611-6595, USA
[2] Department of Radiation Oncology, University of Florida, Gainesville, Florida 32610-0385, USA

E-mail: romeijn@ise.ufl.edu, dempsey@ufl.edu and lijg@ufl.edu

Received 6 January 2004
Published 4 May 2004
Online at stacks.iop.org/PMB/49/1991
DOI: 10.1088/0031-9155/49/10/011

Abstract

Models for finding treatment plans for intensity modulated radiation therapy are usually based on a number of structure-based treatment plan evaluation criteria, which are often conflicting. Rather than formulating a model that *a priori* quantifies the trade-offs between these criteria, we consider a multi-criteria optimization approach that aims at finding the so-called *undominated* treatment plans. We present a unifying framework for studying multi-criteria optimization problems for treatment planning that establishes conditions under which treatment plan evaluation criteria can be transformed into convex criteria while preserving the set of undominated treatment plans. Such transformations are identified for many of the criteria that have been proposed to date, establishing equivalences between these criteria. In addition, it is shown that the use of a nonconvex criterion can often be avoided by transformation to an equivalent convex criterion. In particular, we show that models employing criteria such as tumour control probability, normal tissue complication probability, probability of uncomplicated tumour control, as well as sigmoidal transformations of (generalized) equivalent uniform dose are equivalent to models formulated in terms of separable voxel-based criteria that penalize dose in individual voxels.

1. Introduction

Intensity modulated radiation therapy (IMRT) inverse treatment planning necessitates the formulation of a fluence map optimization (FMO) model to guide the planning process. Such models are generally based on sets of conflicting treatment plan evaluation criteria. For

0031-9155/04/101991+23$30.00

example, one may consider measures of (generalized) equivalent uniform dose (EUD and gEUD), tumour control probability (TCP), normal tissue complication probability (NTCP) and dose- or dose-volume-based criteria. The distinguishing feature of different FMO models then lies in the choice of evaluation criteria that are employed. The typical approach has been to formulate a FMO model that makes an *ad hoc* trade-off by imposing bound constraints on criteria and defining an objective function that optimizes the value of a weighted sum of criteria (see Shepard *et al* (1999) for a review).

An important distinction is the difference between the *importance* of a criterion and its *parameterization*. An example to illustrate this point is the structure-based criterion of generalized equivalent uniform dose as proposed by Kutcher and Burman (1989) and Niemierko (1999). This criterion is parametrized by a power parameter that characterizes the radiation response of the underlying tissue of the structure. Due to the biological meaning that is generally attributed to this parameter, it should be estimated from clinical data, rather than determined in a trial-and-error fashion by repeatedly solving FMO models based on this criterion in an attempt to find clinically valuable treatment plans. Put differently, it must be realized that the relative importance of the gEUD criteria between different structures is not implicit in the values of the corresponding gEUD parameters, but should be measured in another way. For example, one might employ the gEUD of the dose distribution in a spinal cord to evaluate the risk of myelitis while the gEUD of the dose distribution in a saliva gland might be employed to evaluate the risk of xerostomia. In both cases, the gEUD parameter depends on the characteristics of the structure. However, the values of these parameters do not define the relative importance of the risk of myelitis versus the risk of xerostomia, while from a clinical standpoint there is an obvious difference between importance of the two risks. Therefore, it is desirable to incorporate an additional parameter measuring the importance or weight associated with each criterion in a FMO model, which can then be used to control the trade-off between disparate criteria.

Unfortunately, there is generally no scientific or fundamental basis for making a trade-off (i.e., choosing the criteria weights and bounds) between criteria in a given model. A more elegant approach that recognizes this fact is to formulate the FMO model as a multi-criteria optimization problem, as recently proposed by Küfer *et al* (2000), Hamacher and Küfer (2002), Thieke *et al* (2003a, 2003b), Bortfeld *et al* (2003) and Lahanas *et al* (2003). In a multi-criteria approach we are only interested in plans with the property that improving a single criterion value is only possible if at least one other criterion value deteriorates. Treatment plans that possess this property are called *undominated* plans. The multi-criteria approach therefore eliminates all treatment plans for which no trade-off is needed to improve at least one criterion value.

Solving a multi-criteria optimization model now entails characterizing its set of undominated treatment plans. The goal of this paper is to provide a unifying mathematical framework that allows for a scientific comparison of different models via the comparison of the corresponding sets of undominated plans. From a practical point of view, such a comparison can only be made if these sets can be characterized efficiently. As we will see below, the set of undominated plans is convex when only convex treatment plan evaluation criteria are considered. In this case, undominated plans can be identified by solving a family of convex FMO models to global optimality. However, many of the criteria that have been proposed in the literature are nonconvex, leading to nonconvex sets of undominated treatment plans. Our unifying framework establishes conditions under which such criteria can be transformed into convex criteria with an equivalent set of undominated treatment plans. We then apply this framework to identify such transformations for many of the treatment plan evaluation criteria that have been proposed to date.

2. A multi-criteria optimization approach to fluence-map optimization

2.1. Introduction and notation

We consider the general FMO problem for IMRT, where every patient has a set of defined targets and critical structures to consider. The dose received by a voxel is generally assumed to be a linear function of the beamlet intensities (or weights):

$$D_{js}(\vec{x}) = \sum_{i=1}^{n} D_{ijs} x_i \qquad j = 1, \ldots, v_s; \quad s = 1, \ldots, S \tag{1}$$

where D_{ijs} is the dose deposited in voxel j in structure s from beamlet i at unit intensity, and the intensity of beamlet i is denoted by x_i. Furthermore, S is the total number of structures, the first T of which are the targets. Finally, the number of voxels contained in structure s is v_s, and the number of beamlets is n. It will be convenient to denote the dose distribution in a structure as a vector function of the vector of beamlet intensities $\vec{x}$ by

$$\vec{D}_s(\vec{x}) = (D_{1s}(\vec{x}), \ldots, D_{v_s s}(\vec{x})) \qquad s = 1, \ldots, S$$

which can be interpreted as a point in the v_s-dimensional space of all possible dose distributions for structure s.

In the IMRT literature, many alternative types of criteria have been proposed for evaluating treatment plans. Usually, evaluation criteria are formulated for individual structures, and can be expressed as a function of the beamlet intensities $\vec{x}$. To adequately evaluate a treatment plan, we are interested in the values of multiple evaluation criteria (usually at least one per structure). In particular, we assume that we are interested in the values of L criterion functions, denoted by $G_\ell(\vec{x})$ $(\ell = 1, \ldots, L)$. Without loss of generality, we also assume that lower values are preferred to higher values for each of these L criteria (if this is not the case for some criteria, we simply multiply them by -1). Such criteria are generally conflicting, and there will usually not exist a treatment plan that achieves the smallest value of all criteria simultaneously. We therefore need to make a trade-off between these criteria. As mentioned above, there is generally no fundamental basis for making these trade-offs. We therefore formulate the FMO model as a multi-criteria optimization problem, and develop a unifying framework for such models. From this vantage point, we then explore the relationship between multi-criteria FMO models based on different evaluation criteria.

2.2. Multi-criteria optimization

Multi-criteria optimization is an approach that has been widely applied in many traditional engineering and business situations where a trade-off needs to be made between different conflicting criteria (for example, cost and quality in engineering design, and cost and customer service in business; see the textbooks by Steuer (1986) and Miettinen (1999) for a discussion of this general technique). The nature of the FMO problem makes it very suitable for a multi-criteria approach as well. Using the notation introduced in section 2.1, we obtain the following multi-criteria FMO model formulation (P):

$$\text{minimize}_{\vec{x} \geqslant 0} \{G_1(\vec{x}), \ldots, G_L(\vec{x})\}.$$

The concept of *Pareto efficiency* (sometimes also called *Pareto optimality*; see Pareto (1971)) can then be used to characterize all meaningful candidate solutions that should be considered in making the trade-off between the conflicting goals. In particular, solutions of the multi-criteria optimization problem (P) with the property that improving a single criterion value is only possible if at least one other criterion value deteriorates are called Pareto efficient.

On the other hand, solutions for which it is possible to improve the value of one criterion without making the value of any other criterion worse are *inefficient* and therefore not worth considering. As an illustration, suppose that we are given a treatment plan, i.e., a set of beamlet intensities $\vec{x}'$, with corresponding values $G_\ell(\vec{x}')$ for all criteria ($\ell = 1, \ldots, L$). If it is possible to find another treatment plan, say $\vec{x}''$, for which none of the criteria has a higher value while it has a lower value for at least one of the criteria, then the treatment plan $\vec{x}''$ is said to dominate $\vec{x}'$, and we may discard $\vec{x}'$ from further consideration. All treatment plans with the property that it is impossible to find another treatment plan that dominates it are called *Pareto efficient* (or undominated, as defined above) treatment plans.

Before we characterize the Pareto efficient plans in (P) we first define the set

$$B = \{\vec{U} : \exists \vec{x} \geqslant 0 \text{ such that } G_\ell(\vec{x}) \leqslant U_\ell \text{ for } \ell = 1, \ldots, L\}$$

to be the set of vectors of upper bounds $\vec{U} = (U_1, \ldots, U_L)$ for which there exists at least one treatment plan where none of the criteria have a value that exceeds the corresponding upper bound. Note that the set B has the following obvious yet important property.

Property 2.1. *If $\vec{U} \in B$, then if $\vec{U}' \geqslant \vec{U}$, it follows that $\vec{U}' \in B$ as well.*

Now consider the boundary, ∂B, of the set B. This is the set of vectors $\vec{U} \in B$ with the property that no vector $\vec{U}' \in B$ exists for which $\vec{U}' \leqslant \vec{U}$ and $\vec{U}' \neq \vec{U}$ (i.e., $U'_\ell < U_\ell$ for at least one $\ell = 1, \ldots, L$). In other words, a vector of upper bounds $\vec{U} \in \partial B$ has the property that no treatment plan $\vec{x}$ can be found if the bound on one criterion is tightened without relaxing the bound on at least one other criterion. Pareto efficient treatment plans $\vec{x}$ are therefore precisely the treatment plans for which the criteria satisfy a vector of upper bounds on the boundary ∂B of the set B. We call the boundary ∂B the *Pareto efficient frontier* of the set B.

In the following section, we will study how the sets B and ∂B behave under increasing transformations of the treatment plan evaluation criteria $G_\ell(\vec{x})$ ($\ell = 1, \ldots, L$).

2.3. Invariance under increasing transformations of criteria

In this section, we will show that transforming any or all of the criteria used in the multi-criteria FMO model via increasing functions leads to an *equivalent* Pareto efficient frontier. This has very significant consequences for multi-criteria FMO modelling, since it means that we may wisely choose transformations in such a way that the sets B and ∂B take on a mathematical form that is highly efficient from an algorithmic point of view.

In particular, suppose that we transform criterion function G_ℓ by an increasing function h_ℓ (for $\ell = 1, \ldots, L$). We could then consider the associated multi-criteria optimization problem (P^h):

$$\text{minimize}_{\vec{x} \geqslant 0} \{h_1(G_1(\vec{x})), \ldots, h_L(G_L(\vec{x}))\}.$$

In the remainder of this section, we will show that the sets of Pareto efficient treatment plans corresponding to the multi-criteria FMO models (P) and (P^h) are equivalent.

Analogous to the approach in section 2.2, we define the set

$$B^h = \left\{\vec{U}^h : \exists \vec{x} \geqslant 0 \text{ such that } h_\ell(G_\ell(\vec{x})) \leqslant U_\ell^h \text{ for } \ell = 1, \ldots, L\right\}$$

to be the set of vectors of upper bounds $\vec{U}^h = \left(U_1^h, \ldots, U_L^h\right)$ for which there exists at least one treatment plan where none of the transformed criteria have a value that exceeds the corresponding upper bound. Furthermore, we denote the set of Pareto efficient points in B^h (i.e., its boundary) by ∂B^h.

Now observe that the inequalities

$$G_\ell(\vec{x}) \leqslant U_\ell$$

and

$$h_\ell(G_\ell(\vec{x})) \leqslant h_\ell(U_\ell)$$

are mathematically equivalent due to the fact that the function h_ℓ is increasing. Therefore, it is easy to see that there is a one-to-one correspondence between the sets B and B^h:

$$B = \left\{\left(h_1^{-1}(U_1^h), \ldots, h_L^{-1}(U_L^h)\right) : \vec{U}^h \in B^h\right\}$$

and

$$B^h = \{(h_1(U_1), \ldots, h_L(U_L)) : \vec{U} \in B\}. \tag{2}$$

The following theorem contains the first main result of this paper. It shows that there is also a one-to-one correspondence between the set of Pareto efficient points of the sets B and B^h.

Theorem 2.2. *The sets ∂B and ∂B^h are related through*

$$\partial B = \left\{\left(h_1^{-1}(U_1^h), \ldots, h_L^{-1}(U_L^h)\right) : \vec{U}^h \in \partial B^h\right\}$$

or, equivalently,

$$\partial B^h = \{(h_1(U_1), \ldots, h_L(U_L)) : \vec{U} \in \partial B\}.$$

Proof. Suppose $\vec{U} \in \partial B$. We will show that $(h_1(U_1), \ldots, h_L(U_L)) \in \partial B^h$. By equation (2), we have that $(h_1(U_1), \ldots, h_L(U_L)) \in B^h$. Now suppose that

$$(h_1(U_1), \ldots, h_L(U_L)) \notin \partial B^h. \tag{3}$$

Then there exists some $\vec{U}^h \in \partial B^h$ such that

$$U_\ell^h \leqslant h_\ell(U_\ell) \qquad \ell = 1, \ldots, L$$

and strict inequality holds for at least one ℓ. Now consider the transformation of this vector $\left(h_1^{-1}(U_1^h), \ldots, h_L^{-1}(U_L^h)\right) \in B$. It is easy to see that

$$h_\ell^{-1}(U_\ell^h) \leqslant U_\ell \qquad \ell = 1, \ldots, L$$

with strict inequality holding for at least one ℓ. This contradicts the fact that $\vec{U} \in \partial B$, so the assumption in equation (3) is false, and we must instead have that

$$(h_1(U_1), \ldots, h_L(U_L)) \in \partial B^h.$$

A similar argument can be used to show that if $\vec{U}^h \in \partial B^h$, we have that $\left(h_1^{-1}(U_1^h), \ldots, h_L^{-1}(U_L^h)\right) \in \partial B$. This completes the proof. □

2.4. Convex criteria

In this section, we will consider the case where all criterion functions G_ℓ are continuous convex functions of the beamlet intensities $\vec{x}$. In that case, deciding whether some vector of upper bounds $\vec{U}$ is in the set B can be formulated as a convex optimization problem. This is attractive since efficient algorithms exist to solve such problems to *global* optimality. The following proposition shows that in this case the set B is convex and closed.

Proposition 2.3. *If the criterion functions G_ℓ ($\ell = 1, \ldots, L$) are convex and continuous, then the set B is convex and closed.*

Proof. Let $\vec{U}'$ and $\vec{U}''$ be elements of B, and let $\lambda \in [0, 1]$. Let $\vec{x}' \geqslant 0$ be such that $G_\ell(\vec{x}') \leqslant U'_\ell$ for all $\ell = 1, \ldots, L$, and $\vec{x}'' \geqslant 0$ be such that $G_\ell(\vec{x}'') \leqslant U''_\ell$ for all $\ell = 1, \ldots, L$. Clearly, $\lambda\vec{x}' + (1 - \lambda)\vec{x}'' \geqslant 0$. In addition,

$$\begin{aligned} G_\ell(\lambda\vec{x}' + (1 - \lambda)\vec{x}'') &\leqslant \lambda G_\ell(\vec{x}') + (1 - \lambda)G_\ell(\vec{x}'') \\ &\leqslant \lambda U'_\ell + (1 - \lambda)U''_\ell \end{aligned}$$

for $\ell = 1, \ldots, L$ by the convexity of the functions G_ℓ. Therefore, $\lambda\vec{U}' + (1 - \lambda)\vec{U}'' \in B$, and we conclude that B is convex. The fact that B is closed follows directly from the continuity of the functions G_ℓ. □

In the case of convex criteria, we may characterize the Pareto efficient frontier ∂B using a method that is different but equivalent to that used in section 2.2. In particular, we will define a family of optimization problems, the optimal solutions of which yield the Pareto efficient frontier ∂B. Due to the convexity, the closedness and property 2.1, any linear function with non-negative coefficients minimized over the set B yields a point on the boundary (i.e., the efficient frontier) ∂B. Using the definition of the set B, this means that the efficient frontier ∂B can be found by solving the following optimization problem (P($\vec{\mu}$)):

$$\text{minimize} \sum_{\ell=1}^{L} \mu_\ell U_\ell$$

subject to

$$\begin{aligned} & G_\ell(\vec{x}) \leqslant U_\ell \qquad \ell = 1, \ldots, L \\ & \vec{x} \geqslant 0 \\ & U_\ell \text{ free} \qquad \ell = 1, \ldots, L \end{aligned} \tag{4}$$

for all parameter vectors $\vec{\mu} \geqslant 0$. In other words, for each $\vec{U}' \in \partial B$, there exists a treatment plan $\vec{x}' \geqslant 0$ and a parameter vector $\vec{\mu} \geqslant 0$ such that $(\vec{x}', \vec{U}')$ is an optimal solution to the optimization problem (P($\vec{\mu}$)). The following theorem contains the second main result of this paper by providing a simplification of the formulation of this problem.

Theorem 2.4. *The optimization problem* (P($\vec{\mu}$)) *can be equivalently formulated as*

$$\textit{minimize}_{\vec{x} \geqslant 0} \sum_{\ell=1}^{L} \mu_\ell G_\ell(\vec{x})$$

Proof. We will show that, without loss of optimality, we may replace the inequality constraints (4) in (P($\vec{\mu}$)) by equality constraints. Suppose that this statement is untrue, i.e., suppose that $(\vec{x}^*, \vec{U}^*)$ is an optimal solution to (P($\vec{\mu}$)), and one of the criterion bound constraints (4) is satisfied as a strict inequality: $G_{\hat{\ell}}(\vec{x}^*) < U_{\hat{\ell}}$ for some $\hat{\ell}$. Then the solution $(\vec{x}^*, \vec{U}')$ with

$$\begin{aligned} & U'_{\hat{\ell}} = G_{\hat{\ell}}(\vec{x}^*) \\ & U'_\ell = U^*_\ell \qquad \text{for} \quad \ell = 1, \ldots, L; \ell \neq \hat{\ell} \end{aligned}$$

is a feasible solution to (P($\vec{\mu}$)) as well, and its solution value is no worse than the solution value of $(\vec{x}^*, \vec{U}^*)$ since $\mu_{\hat{\ell}} \geqslant 0$. Therefore, $(\vec{x}^*, \vec{U}')$ is an optimal solution to (P($\vec{\mu}$)) as well. Repeating this argument for all bound constraints that are satisfied as a strict inequality we arrive at the desired result. The equivalent formulation of the optimization problem (P($\vec{\mu}$)) then follows by substituting the criterion functions $G_\ell(\vec{x})$ for the upper bounds U_ℓ. □

Clearly, the problem (P($\vec{\mu}$)) is a convex optimization problem by the convexity of the functions G_ℓ ($\ell = 1, \ldots, L$). In summary, we have shown that if the criterion functions are

all convex, we can find all solutions on the efficient frontier ∂B with respect to these criteria by solving convex optimization problems of the form (P($\vec{\mu}$)), where the coefficients $\vec{\mu}$ can be viewed as relative importance factors assigned to the individual criteria.

3. Relationships between treatment plan evaluation criteria

3.1. A unifying framework for treatment plan evaluation criteria

In this section, we will show that most structure-based treatment plan evaluation criteria that have been proposed in the literature can be cast in the form

$$(h \circ G)(\vec{x}) \equiv h(G(\vec{x})) \tag{5}$$

where G is a convex criterion for which small values are preferred to large values and where h is an increasing function. Compositions of this type provide a unifying framework for studying the relationships between various treatment plan evaluation criteria.

If we have specified a set of composite criteria $h_\ell \circ G_\ell$ ($\ell = 1, \ldots, L$) as described above, theorem 2.2 states that the Pareto efficient frontier with respect to the criteria $\{G_\ell\}$ is equivalent to that with respect to the criteria $\{h_\ell \circ G_\ell\}$. *This means that the set of Pareto efficient treatment plans is the same for both choices of criteria.* Moreover, theorem 2.4 states that we can find these Pareto efficient plans by solving convex optimization problems of the form (P($\vec{\mu}$)). Clearly, theorem 2.4 further states that if the criteria $\{h_\ell \circ G_\ell\}$ are convex, we can also find the Pareto efficient plans by solving the following alternative family of convex optimization problems (P^h($\vec{\mu}^h$)):

$$\text{minimize}_{\vec{x} \geqslant 0} \sum_{\ell=1}^{L} \mu_\ell^h h_\ell(G_\ell(\vec{x}))$$

for $\vec{\mu}^h \geqslant 0$. This is relevant if the optimization problems (P^h($\vec{\mu}^h$)) happen to be more efficiently solvable than (P($\vec{\mu}$)). Perhaps more importantly, if the criteria $\{h_\ell \circ G_\ell\}$ are *not* convex, theorem 2.4 then states that the Pareto efficient treatment plans with respect to these criteria can still be found by solving the optimization problems (P($\vec{\mu}$)), which are formulated with respect to the convex criteria $\{G_\ell\}$. *This allows one to replace a nonconvex criterion by an equivalent convex one if a suitable decomposition into a convex criterion G and an increasing function h can be found.*

3.2. Criteria as functions of dose versus beamlet intensities

In the literature, structure-based treatment plan evaluation criteria have usually been cast as a function of the dose distribution in a structure, rather than as a function of the beamlet intensities $\vec{x}$. With a slight abuse of notation, we will use the same symbol, G, for both. That is, if $\vec{d}$ denotes the dose distribution in a given structure, we will use both $G(\vec{x})$ and $G(\vec{d})$ to evaluate a treatment plan. Although the functions $G(\vec{x})$ and $G(\vec{d})$ are formally speaking different, the interpretation of the argument will be clear from the context. Moreover, the two representations are related through

$$G(\vec{x}) = G(\vec{D}(\vec{x}))$$

using the dose model in equation (1) and suppressing, as we will do whenever appropriate in the remainder of this section, the subscript s indicating the particular structure to which the criterion is applied.

In order to study treatment plan evaluation criteria and cast them in terms of our framework (5), it is important to be able to determine whether a given criterion is a convex function of

the beamlet intensities. As the following lemma shows, convexity of an evaluation criterion in the dose distribution implies that it is convex in the beamlet intensities as well under the linear dose model (1), so that we may limit ourselves to studying the convexity of treatment plan evaluation criteria as a function of the dose distribution.

Lemma 3.1. *If $G(\vec{d})$ is a convex function of $\vec{d}$, then $G(\vec{x}) = G(\vec{D}(\vec{x}))$ is a convex function of $\vec{x}$.*

Proof. Let $\vec{x}', \vec{x}'' \geqslant 0$ be two vectors of beamlet intensities, and let $\lambda \in [0, 1]$. Since $\vec{D}(\vec{x})$ is a linear function of $\vec{x}$, we have

$$\begin{aligned} G(\lambda\vec{x}' + (1-\lambda)\vec{x}'') &= G(\vec{D}(\lambda\vec{x}' + (1-\lambda)\vec{x}'')) \\ &= G(\lambda\vec{D}(\vec{x}') + (1-\lambda)\vec{D}(\vec{x}'')) \\ &\leqslant \lambda G(\vec{D}(\vec{x}')) + (1-\lambda)G(\vec{D}(\vec{x}'')) \\ &= \lambda G(\vec{x}') + (1-\lambda)G(\vec{x}'') \end{aligned}$$

which proves that $G(\vec{x})$ is a convex function of $\vec{x}$. □

In the following subsection, we will analyse many proposed treatment plan evaluation criteria from the literature and study their relationships via our unifying framework (5).

3.3. Analysis of commonly used criteria

3.3.1. Tumour control probability. The response of a target to irradiation can be characterized by the probability that no clonogen cells remain in the target, often called the *tumour control probability* (see, e.g., Withers and McBride (1998)). Based on a Poisson approximation of a binomial model for the number of remaining clonogens and including cell repopulation effects, the TCP of a target can be expressed as

$$\prod_{j=1}^{v} \exp\left(-\frac{N\,\mathrm{e}^{\lambda(n-1)\Delta T}}{v} \cdot p(d_j)\right) \tag{6}$$

where N is the total number of clonogen cells in the target, $p(d_j)$ is the surviving fraction of clonogen cells in voxel j of the target after receiving dose d_j, n is the number of fractions used for the treatment, ΔT is the time between two consecutive fractions and λ is the net rate of cell birth (see Stavreva *et al* (2003) and Zaider and Minerbo (2000)). The recent work by Stavreva *et al* (2003) shows that a single-hit model with repopulation provides the best empirical fit to tumour response data. In this case, the TCP for a target becomes

$$\mathrm{TCP}(\vec{d}; N, \alpha, \lambda, n, \Delta T) = \prod_{j=1}^{v} \exp\left(-\frac{N\,\mathrm{e}^{\lambda(n-1)\Delta T}}{v} \cdot \mathrm{e}^{-\alpha d_j}\right) \tag{7}$$

where $1/\alpha$ is the mean lethal dose to the target, and describes the radioresistance of cells in that target. The single-hit model without repopulation (see, e.g., Brahme and Agren (1987) and Goitein (1987)) is obtained as a special case by substituting $\lambda = 0$:

$$\mathrm{TCP}(\vec{d}; N, \alpha) = \prod_{j=1}^{v} \exp\left(-\frac{N}{v} \cdot \mathrm{e}^{-\alpha d_j}\right).$$

3.3.2. Equivalent uniform dose. Niemierko (1997) introduced the concept of *equivalent uniform dose* for a target by equating the TCP for an inhomogeneous dose distribution to that for a homogeneous dose distribution and solving for the corresponding homogeneous dose:

$$\mathrm{TCP}(\vec{d}) = \mathrm{TCP}(\mathrm{EUD}\vec{1})$$

where $\vec{1}$ is a vector in which all elements are equal to 1. Applying this approach to the model for the TCP of a target given in equation (7), we obtain Niemierko's original definition of EUD by setting

$$\mathrm{TCP}(\vec{d};\, N, \alpha, \lambda, n, \Delta T) = \prod_{j=1}^{v} \exp\left(-\frac{N\,\mathrm{e}^{\lambda(n-1)\Delta T}}{v} \cdot \mathrm{e}^{-\alpha d_j}\right)$$

$$= \exp\left(-N\,\mathrm{e}^{\lambda(n-1)\Delta T} \cdot \frac{1}{v}\sum_{j=1}^{v} \mathrm{e}^{-\alpha d_j}\right)$$

equal to

$$\mathrm{TCP}(\mathrm{EUD}\vec{1};\, N, \alpha, \lambda, n, \Delta T) = \exp(-N\,\mathrm{e}^{\lambda(n-1)\Delta T} \cdot \mathrm{e}^{-\alpha \mathrm{EUD}(\vec{d})})$$

yielding

$$\mathrm{EUD}(\vec{d};\, \alpha) \equiv -\frac{1}{\alpha}\ln\left(\frac{1}{v}\sum_{j=1}^{v} \mathrm{e}^{-\alpha d_j}\right)$$

(see also McGary *et al* (2000)). Note that this measure of EUD is independent of the repopulation rate λ, the number of fractions n, and the time between fractions ΔT.

Subsequently, Niemierko (1999) (using a discrete version of a model earlier proposed by Kutcher and Burman (1989)), defined the concept of *generalized EUD* for a target as the generalized mean of the dose distribution in that target, i.e.,

$$\mathrm{gEUD}(\vec{d};\, a) = \left(\frac{1}{v}\sum_{j=1}^{v} d_j^a\right)^{1/a}$$

(see also Abramowitz and Stegun (1965)), where $-\infty \leqslant a \leqslant 0$ is a structure-dependent parameter that depends on the radiation response of the underlying tissue. For parameter values $1 \leqslant a \leqslant \infty$, this definition extends the concept of gEUD to critical structures. Note that for the limiting cases $a = -\infty$, 0 and ∞, the gEUD becomes the minimum, geometric mean and maximum dose in the structure, respectively. Since its introduction, it has been used by many researchers to formulate apparently different models for FMO; for examples please see Choi and Deasy (2002), Wu *et al* (2002), Thieke *et al* (2003c), and Bortfeld *et al* (2003). In the following section, we will study one of these models in greater detail.

3.3.3. Sigmoidal criteria based on gEUD. Wu *et al* (2002) have proposed a model for FMO that summarizes the dose distribution in each structure by a function of the gEUD. In particular, their model for FMO involves the maximization of the following objective function:

$$\prod_{s=1}^{S} w_s\left(\vec{d}_s;\, a_s, k_s, \mathrm{gEUD}_s^0\right)$$

or, equivalently, the minimization of

$$-\sum_{s=1}^{S} \ln w_s\left(\vec{d}_s;\, a_s, k_s, \mathrm{gEUD}_s^0\right)$$

 8717

where a_s is the gEUD parameter, gEUD_s^0 is a base gEUD value, and k_s is a positive parameter ($s = 1, \ldots, S$). The functions w_s are chosen to be logistic functions, which have a sigmoidal shape. In particular, for a generic structure

$$w(\vec{d}; a, k, \text{gEUD}^0) = \begin{cases} \dfrac{1}{1 + \left(\frac{\text{gEUD}^0}{\text{gEUD}(\vec{d}; a)}\right)^k} & \text{if} \quad -\infty \leqslant a \leqslant 0 \\ \dfrac{1}{1 + \left(\frac{\text{gEUD}(\vec{d}; a)}{\text{gEUD}^0}\right)^k} & \text{if} \quad 1 \leqslant a \leqslant \infty. \end{cases}$$

Note that gEUD^0 is in fact the inflection point of the sigmoidal function, and k determines the steepness of the sigmoidal function at the inflection point. In a multi-criteria setting, this suggests the use of the following treatment plan evaluation criteria for structures:

$$W(\vec{d}; a, k, \text{gEUD}^0) = -\ln w(\vec{d}; a, k, \text{gEUD}^0) = \begin{cases} \ln\left(1 + \left(\dfrac{\text{gEUD}^0}{\text{gEUD}(\vec{d}; a)}\right)^k\right) & \text{if} \quad -\infty \leqslant a \leqslant 0 \\ \ln\left(1 + \left(\dfrac{\text{gEUD}(\vec{d}; a)}{\text{gEUD}^0}\right)^k\right) & \text{if} \quad 1 \leqslant a \leqslant \infty. \end{cases}$$

As Wu *et al* (2002) mention, they could have chosen another sigmoidal function of gEUD than the logistic one to define their criteria. We will therefore also study what the consequences are of employing Φ, the cumulative distribution function of the standard normal distribution (i.e., $\Phi(z) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{z} e^{-t^2/2}\, dt$). In that case, we obtain the criteria

$$\breve{W}(\vec{d}; a, \sigma, \text{gEUD}^0) = -\ln \breve{w}(\vec{d}; a, \sigma, \text{gEUD}^0) = \begin{cases} -\ln\left(1 - \Phi\left(\dfrac{\text{gEUD}^0 - \text{gEUD}(\vec{d}; a)}{\sigma \cdot \text{gEUD}^0}\right)\right) & \text{if} \quad -\infty \leqslant a \leqslant 0 \\ -\ln\left(1 - \Phi\left(\dfrac{\text{gEUD}(\vec{d}; a) - \text{gEUD}^0}{\sigma \cdot \text{gEUD}^0}\right)\right) & \text{if} \quad 1 \leqslant a \leqslant \infty. \end{cases}$$

Again, gEUD^0 is the inflection point of the sigmoidal function; however, in this case the parameter $\sigma > 0$ determines the steepness of the sigmoidal function at the inflection point.

3.3.4. Normal tissue complication probability. Several models have been proposed in the literature to describe the *normal tissue complication probability* associated with the dose distribution in a critical structure (see, e.g., Wolbarst (1984), Lyman (1985), Lyman and Wolbarst (1987a, 1987b), Schultheiss *et al* (1983), Niemierko and Goitein (1991)). In this section, we will discuss the most recent work by Stavrev *et al* (2003) and Alber and Nüsslin (2001).

Stavrev *et al* (2003) base the NTCP for inhomogeneous dose distributions on the following model that Lyman (1985) proposed for determining the NTCP for a homogeneous dose distribution in a critical structure:

$$\text{NTCP}(D\vec{i}; m, D_{50}) = \Phi\left(\frac{D - D_{50}}{m D_{50}}\right) \tag{8}$$

where D is the homogeneous dose, D_{50} denotes the uniform dose where the structure exhibits a 50% complication probability, and the parameter m determines the mid-point

slope of the NTCP curve, sometimes also called γ_{50}. In particular, they determine the NTCP for inhomogeneous dose distributions by evaluating equation (8) at the gEUD for the inhomogeneous dose distribution:

$$\begin{aligned}\mathrm{NTCP}^{\mathrm{L\&S}}(\vec{d}; a, m, D_{50}) &= \mathrm{NTCP}(\mathrm{gEUD}(\vec{d}; a)\vec{1}; m, D_{50}) \\ &= \Phi\left(\frac{\mathrm{gEUD}(\vec{d}; a) - D_{50}}{m D_{50}}\right)\end{aligned}$$

where $1 \leqslant a \leqslant \infty$ is the gEUD parameter associated with the critical structure.

Alber and Nüsslin (2001) have proposed a phenomenological description of an NTCP function derived from mechanistic concepts. They arrive at the expression

$$\mathrm{NTCP}^{\mathrm{A\&N}}(\vec{d}; a, \Delta) = 1 - \exp\left(-\left(\frac{\mathrm{gEUD}(\vec{d}; a)}{\Delta}\right)^{a}\right)$$

where again $1 \leqslant a < \infty$ is the gEUD parameter associated with the critical structure and Δ is a structure-dependent constant denoting the uniform dose where the structure exhibits a $1 - \mathrm{e}^{-1} \approx 63\%$ complication probability.

3.3.5. Separable convex voxel-based criteria. The treatment plan evaluation criteria described thus far are all based on attempts at modelling the biological effect of irradiating the tissues underlying the contoured structures in the FMO problem. However, an approach that has been more widely used to date is one that heuristically measures treatment plan quality by evaluating the dose received by each voxel through a convex function. By considering the average measure received by all voxels in a structure, we can express such models in the context of this paper by defining a treatment plan evaluation criterion for a structure that is *separable* in the doses to individual voxels

$$F(\vec{d}) = \frac{1}{v}\sum_{j=1}^{v} f_j(d_j)$$

where f_j is a convex function. The most common choices found in the literature are of the form

$$f_j(d_j) = c_j|d_j - \delta_j|^{a_j}$$

for targets, and

$$f_j(d_j) = c_j \max\{0, d_j - \delta_j\}^{a_j}$$

for critical structures. Here, δ_j is a dose threshold and a_j and c_j are shape and scale parameters for the convex function corresponding to voxel j in the structure. The shape parameter a_j is usually chosen to be 1 or 2 (where $a_j = 2$ corresponds to the often used least squares model for FMO), and the parameters δ_j and c_j usually depend only on the structure (see Shepard *et al* (1999) for a review of such models). Recently, Tsien *et al* (2003) and Romeijn *et al* (2003) have obtained excellent results by using high powers of dose difference or a piecewise-linear approximation thereof, respectively. Other choices that will prove useful in the remainder of this section are obtained when choosing

$$f_j(d_j; \alpha) = \mathrm{e}^{-\alpha d_j} \qquad \text{for} \quad \alpha > 0$$

or

$$f_j^{\mathrm{g}}(d_j; a) = \begin{cases} d_j^a & \text{for} \quad a \neq 0 \\ -\ln d_j & \text{for} \quad a = 0. \end{cases}$$

We will refer to the corresponding structure-based treatment plan evaluation functions as $F(\vec{d}; \alpha)$ and $F^{\mathrm{g}}(\vec{d}; a)$, respectively.

Table 1. Evaluation criteria for targets.

		Criterion decomposition	
Criterion	*Parameter range*	$G(\vec{d})$	$h(z)$
$\mathrm{TCP}(N, \alpha, \lambda, n, \Delta T)$	$N, \alpha > 0; n \geqslant 1; \lambda, \Delta T \geqslant 0$	$-\ln \mathrm{TCP}(\vec{d}; N, \alpha, n, \Delta T, \lambda)$	$-e^{-z}$
		$-\mathrm{EUD}(\vec{d}; \alpha)$	$-\exp(-Ne^{\lambda(n-1)\Delta T+\alpha z})$
$\mathrm{EUD}(\alpha)$	$\alpha > 0$	$-\mathrm{EUD}(\vec{d}; \alpha)$	z
		$F(\vec{d}; \alpha)$	$\frac{1}{\alpha} \ln z$
$\mathrm{gEUD}(a)$	$-\infty \leqslant a \leqslant 0$	$-\mathrm{gEUD}(\vec{d}; a)$	z
	$-\infty < a < 0$	$F^{\mathrm{g}}(\vec{d}; a)$	$-z^{1/a}$
	$a = 0$	$F^{\mathrm{g}}(\vec{d}; 0)$	$-e^{-z}$
$W(a, k, \mathrm{gEUD}^0)$	$-\infty \leqslant a \leqslant 0; k, \mathrm{gEUD}^0 > 0$	$W(\vec{d}; a, k, \mathrm{gEUD}^0)$	z
		$-\mathrm{gEUD}(\vec{d}; a)$	$\ln\left(1 + \left(-\frac{\mathrm{gEUD}^0}{z}\right)^k\right)$
$\tilde{W}(a, \sigma, \mathrm{gEUD}^0)$	$-\infty \leqslant a \leqslant 0; \sigma, \mathrm{gEUD}^0 > 0$	$\tilde{W}(\vec{d}; a, \sigma, \mathrm{gEUD}^0)$	z
		$-\mathrm{gEUD}(\vec{d}; a)$	$-\ln\left(1 - \Phi\left(\frac{\mathrm{gEUD}^0+z}{\sigma \cdot \mathrm{gEUD}^0}\right)\right)$
$F(\alpha)$	$\alpha > 0$	$F(\vec{d}; \alpha)$	z
	$\alpha > 0$	$-\mathrm{EUD}(\vec{d}; \alpha)$	e^z
$F^{\mathrm{g}}(a)$	$-\infty < a \leqslant 0$	$F^{\mathrm{g}}(\vec{d}; a)$	z
	$-\infty < a < 0$	$-\mathrm{gEUD}(\vec{d}; a)$	$(-z)^a$

Table 2. Evaluation criteria for critical structures.

Criterion	*Parameter range*	$G(\vec{d})$	$h(z)$
$\mathrm{gEUD}(a)$	$1 \leqslant a \leqslant \infty$	$\mathrm{gEUD}(\vec{d}; a)$	z
	$1 \leqslant a < \infty$	$F^{\mathrm{g}}(\vec{d}; a)$	$z^{1/a}$
$W(a, k, \mathrm{gEUD}^0)$	$1 \leqslant a \leqslant \infty; k, \mathrm{gEUD}^0 > 0$	$\mathrm{gEUD}(\vec{d}; a)$	$\ln\left(1 + \left(\frac{z}{\mathrm{gEUD}^0}\right)^k\right)$
$\tilde{W}(a, \sigma, \mathrm{gEUD}^0)$	$1 \leqslant a \leqslant \infty; \sigma, \mathrm{gEUD}^0 > 0$	$\mathrm{gEUD}(\vec{d}; a)$	$-\ln\left(1 - \Phi\left(\frac{z-\mathrm{gEUD}^0}{\sigma \cdot \mathrm{gEUD}^0}\right)\right)$
$\mathrm{NTCP}^{\mathrm{L\&S}}(a, m, D_{50})$	$1 \leqslant a \leqslant \infty; m, D_{50} > 0$	$\mathrm{gEUD}(\vec{d}; a)$	$\Phi\left(\frac{z-D_{50}}{mD_{50}}\right)$
$\mathrm{NTCP}^{\mathrm{A\&N}}(a, \Delta)$	$1 \leqslant a < \infty, \Delta > 0$	$\mathrm{gEUD}(\vec{d}; a)$	$1 - e^{-(z/\Delta)^a}$
$F^{\mathrm{g}}(a)$	$1 \leqslant a \leqslant \infty$	$F^{\mathrm{g}}(\vec{d}; a)$	z
	$1 \leqslant a < \infty$	$\mathrm{gEUD}(\vec{d}; a)$	z^a

3.4. Application of the unifying framework

In this section, we apply the unifying framework introduced in section 3.1 to study the relationship between the criteria discussed in section 3.3. To this end, we provide one or more decompositions of each criterion into a convex criterion G and an increasing function h, as required by the framework. Using these decompositions, we can identify the equivalence of Pareto frontiers obtained with different multi-criteria models for FMO.

Some possible decompositions for each of the target criteria are shown in table 1, while table 2 shows decompositions for each of the critical structure criteria. Note that, in section 2 of this paper, we have used the convention that, for all treatment plan evaluation criteria G, smaller values are preferred to larger values. However, for the TCP, EUD and gEUD target criteria in table 1, the opposite is true. Therefore, in each of these cases, we have in fact decomposed the negative of the corresponding criterion.

To verify the validity of the decompositions, we need to establish that the criteria G in the tables are all convex, and that the functions h are all increasing. With respect to the criteria G, we note the following:

- Choi and Deasy (2002) have shown that $\ln \mathrm{TCP}(\vec{d}; N, \alpha)$ is a concave function of $\vec{d}$ when $N, \alpha > 0$; this immediately implies that
$$\ln \mathrm{TCP}(\vec{d}; N, \alpha, n, \Delta T, \lambda) = \mathrm{e}^{\lambda(n-1)\Delta T} \cdot \ln \mathrm{TCP}(\vec{d}; N, \alpha)$$
is a concave function of $\vec{d}$ when $N, \alpha > 0, n \geqslant 1, \Delta T, \lambda \geqslant 0$ as well;
- in appendix A we show that $\mathrm{EUD}(\vec{d}; \alpha)$ is a concave function of $\vec{d}$ when $\alpha > 0$;
- Choi and Deasy (2002) have shown that $\mathrm{gEUD}(\vec{d}; a)$ is a convex function of $\vec{d}$ when $1 \leqslant a \leqslant \infty$, and a concave function of $\vec{d}$ when $-\infty \leqslant a \leqslant 0$;
- in appendix B we show that $W(\vec{d}; a, k, \mathrm{gEUD}^0)$ is a convex function of $\vec{d}$ when $-\infty \leqslant a \leqslant 0$ and $k, \mathrm{gEUD}^0 > 0$.
- in appendix B we also show that $\check{W}(\vec{d}; a, \sigma, \mathrm{gEUD}^0)$ is a convex function of $\vec{d}$ when $-\infty \leqslant a \leqslant 0$ or $1 \leqslant a \leqslant \infty$ and $\sigma, \mathrm{gEUD}^0 > 0$.

Recalling that multiplying a concave function by -1 yields a convex function, the sum of convex functions is convex, the functions $\mathrm{e}^{-\alpha z}$ ($\alpha > 0$) and z^a ($a < 0$ or $a \geqslant 1$ and $z \geqslant 0$) are convex, and the function $\ln z$ is concave, the convexity of all criteria G in tables 1 and 2 now follows. The fact that all functions h in these tables are increasing is easily verified in most cases; for the functions h used in the decomposition of criteria W and $\check{W}$ this can be found in appendix B.

The results of section 2 now show that the efficient frontier with respect to *any* combination of criteria listed in the first column of tables 1 and 2 is *equivalent* to the efficient frontier that is obtained when one or more of these are replaced by a corresponding criterion G from the third column of these tables. This important observation can be summarized as follows:

Observation 3.2. *For any combination of criteria listed in the first column of tables 1 and 2, with the exception of the minimum and maximum dose criteria, there exists a set of criteria that are separable in the voxels that yield an equivalent efficient frontier.*

A previous indication that there exists a close relationship between voxel-based penalty functions and biological treatment plan evaluation criteria appeared in the paper by Alber and Nüsslin (1999), who showed how relaxing TCP and NTCP constraints lead to voxel-based penalty formulations of the FMO problem. In the following section, we will illustrate observation 3.2 by applying the unifying framework to the model by Wu *et al* (2002).

3.5. Example of applying the unifying framework

Wu *et al* (2002) have proposed the following model for FMO:

$$\text{minimize}_{\vec{x} \geqslant 0} \sum_{s=1}^{T} W_s\left(\vec{x}; \tilde{a}_s, \tilde{k}_s, \mathrm{gE\tilde{U}D}_s^0\right) + \sum_{s=1}^{S} W_s\left(\vec{x}; a_s, k_s, \mathrm{gEUD}_s^0\right)$$

where, in addition to the target criteria, they have also applied critical structure criteria to the targets to achieve homogeneity of the target dose distributions. In particular, for target s ($s = 1, \ldots, T$), the parameters a_s, k_s and gEUD_s^0 apply to the 'critical structure nature' of the target, while the parameters $\tilde{a}_s$, $\tilde{k}_s$ and $\mathrm{gE\tilde{U}D}_s^0$ apply to the 'target nature' of the target. Note also that we have returned to expressing the criteria as functions of the beamlet intensities $\vec{x}$ rather than the dose distributions $\vec{d}_s$ (see also section 3.2).

The underlying goal of this model is to simultaneously minimize the values of all criteria W_s. However, there appears to be no fundamental basis for the particular trade-off between these conflicting criteria that they have chosen. We may therefore consider the following more general multi-criteria optimization model, which we will call (P^W):

$$\text{minimize}_{\vec{x}\geqslant 0} \{W_1(\vec{x}; \tilde{a}_1, \tilde{k}_1, \text{gE}\tilde{\text{U}}\text{D}_1^0), \ldots, W_T(\vec{x}; \tilde{a}_T, \tilde{k}_T, \text{gE}\tilde{\text{U}}\text{D}_T^0), W_1(\vec{x}; a_1, k_1, \text{gEUD}_1^0), \ldots, W_S(\vec{x}; a_S, k_S, \text{gEUD}_S^0)\}.$$

The main disadvantage of this model is the fact that the criterion functions are not all convex, as shown in appendix B. This has two severe consequences. Firstly, this implies that, although the globally optimal solutions to problems ($P^W(\vec{\mu}^W, \tilde{\vec{\mu}}^W)$) of the form:

$$\text{minimize}_{\vec{x}\geqslant 0} \sum_{s=1}^{T} \tilde{\mu}_s^W W_s(\vec{x}; \tilde{a}_s, \tilde{k}_s, \text{gE}\tilde{\text{U}}\text{D}_s^0) + \sum_{s=1}^{S} \mu_s^W W_s(\vec{x}; a_s, k_s, \text{gEUD}_s^0)$$

with $\vec{\mu}^W, \tilde{\vec{\mu}}^W \geqslant 0$ are Pareto efficient, they do not characterize the entire Pareto efficient frontier of the model. Secondly, the latter optimization problems are nonconvex, so that efficient methods do not exist for solving these problems to global optimality. Moreover, locally optimal solutions (which can be found relatively easily using local search algorithms) are not necessarily Pareto efficient. (Interestingly, note that if Wu *et al* (2002) had chosen to use the criteria $\check{W}$ instead of W, their objective function would have been convex and therefore unimodal! Moreover, convex optimization problems of the form ($P^{\check{W}}(\vec{\mu}^{\check{W}}, \tilde{\vec{\mu}}^{\check{W}})$) could have been used to characterize the entire Pareto efficient frontier of the corresponding model ($P^{\check{W}}$).) However, if we now apply our unifying framework to (P^W) (or ($P^{\check{W}}$)), we can use tables 1 and 2 to conclude that the Pareto efficient frontier of the following multi-criteria model (P^{gEUD})

$$\text{minimize}_{\vec{x}\geqslant 0} \{-\text{gEUD}_1(\vec{x}; \tilde{a}_1), \ldots, -\text{gEUD}_T(\vec{x}; \tilde{a}_T), \text{gEUD}_1(\vec{x}; a_1), \ldots, \text{gEUD}_S(\vec{x}; a_S)\}$$

is equivalent to that of (P^W). Since the criteria used in (P^{gEUD}) (which is in fact the model proposed by Bortfeld *et al* (2003)) are convex, not only do we know that the Pareto efficient frontier is characterized entirely by the globally optimal solutions to problems of the form

$$\text{minimize}_{\vec{x}\geqslant 0} -\sum_{s=1}^{T} \tilde{\mu}_s^{\text{gEUD}} \text{gEUD}_s(\vec{x}; \tilde{a}_s) + \sum_{s=1}^{S} \mu_s^{\text{gEUD}} \text{gEUD}_s(\vec{x}; a_s)$$

with $\vec{\mu}^{\text{gEUD}}, \tilde{\vec{\mu}}^{\text{gEUD}} \geqslant 0$, but, in addition, the globally optimal solutions to these optimization problems can be found efficiently due to their convexity. It is interesting to note that the parameters k_s, $\tilde{k}_s$, EUD_s^0 and $\text{E}\tilde{\text{U}}\text{D}_s^0$ of (P^W) do not appear in the equivalent model (P^{gEUD}). Clearly, this means that these parameters do not influence the Pareto efficient frontier and are therefore superfluous in a multi-criteria setting. In terms of our unifying framework, any parametrized increasing function h could be applied to the gEUD criteria (for example, the function leading to the criteria $\check{W}$). However, the additional complexity introduced by the corresponding parameters does not change the essence of the model since the set of Pareto efficient solutions is invariant under such transformations.

If all gEUD parameters a_s and $\tilde{a}_s$ are finite, we can now take the application of our unifying framework one step further, and use tables 1 and 2 to apply another set of transformations to arrive at the equivalent multi-criteria optimization model (P^F):

$$\text{minimize}_{\vec{x}\geqslant 0} \{F_1(\vec{x}; \tilde{a}_1), \ldots, F_T(\vec{x}; \tilde{a}_T), F_1(\vec{x}; a_1), \ldots, F_S(\vec{x}; a_S)\}.$$

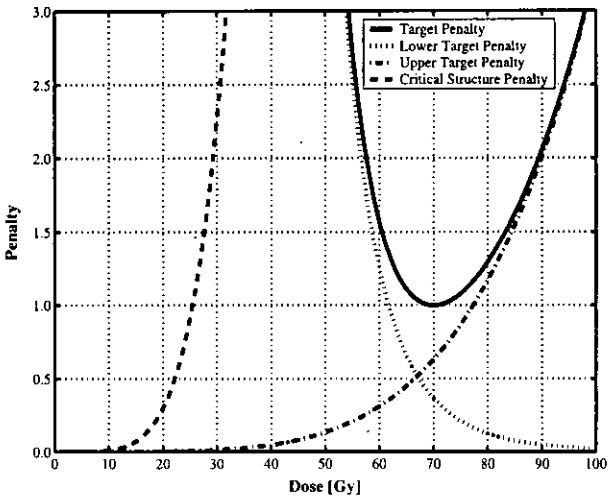


Figure 1. Voxel-based penalty functions for a target and a critical structure.

This result demonstrates that the equivalent Pareto efficient frontiers of (P^W), (P^{gEUD}) and (P^F) can be found by solving optimization problems of the form

$$\text{minimize}_{\vec{x}\geqslant 0} \sum_{s=1}^{T} \tilde{\mu}_s^F F_s(\vec{x}; \tilde{a}_s) + \sum_{s=1}^{S} \mu_s^F F_s(\vec{x}; a_s)$$

with $\vec{\mu}^F, \vec{\tilde{\mu}}^F \geqslant 0$. Note that the objective function of this model can be rewritten as

$$\text{minimize}_{\vec{x}\geqslant 0} \sum_{s=1}^{T} \left(\tilde{\mu}_s^F F_s(\vec{x}; \tilde{a}_s) + \mu_s^F F_s(\vec{x}; a_s)\right) + \sum_{s=T+1}^{S} \mu_s^F F_s(\vec{x}; a_s)$$

or even

$$\text{minimize}_{\vec{x}\geqslant 0} \sum_{s=1}^{T} \sum_{j=1}^{v_s} \left(\tilde{\mu}_s^F f_{js}(\vec{x}; \tilde{a}_s) + \mu_s^F f_{js}(\vec{x}; a_s)\right) + \sum_{s=T+1}^{S} \sum_{j=1}^{v_s} \mu_s^F f_{js}(\vec{x}; a_s).$$

This is a voxel-based penalty function formulation for FMO where all voxels in a given structure have identical penalty functions. The advantage of this formulation over a model with structure-based criteria is the separability of the objective function in the individual voxel-doses, which is often computationally more efficient.

Figure 1 shows an example of voxel-based penalty functions for a target and a critical structure that lead to a multi-criteria optimization model (P^F) that is equivalent to corresponding models (P^W) and (P^{gEUD}). In particular, we follow the gEUD parameter choices made by Wu *et al* (2002) and show a voxel-based penalty function for a head-and-neck target with parameters $\tilde{a} = -8$ and $a = 4.6$

$$f_{js}(d) = \tilde{\mu}_s^F d^{-8} + \mu_s^F d^{4.6}$$

as well as a voxel-based penalty function for a parotid with parameter $a = 5$

$$f_{js}(d) = \mu_s^F d^5$$

where the coefficients $\tilde{\mu}_s^F$ and μ_s^F were chosen for illustrative purposes.

Note that if some of the gEUD parameters are infinite, which means that the minimum or maximum dose in certain structures are in the model (P^{gEUD}), we can incorporate these in the model (P^F) without transforming them. In the corresponding optimization problem ($P^F(\vec{\mu}^F, \vec{\bar{\mu}}^F)$), these can be handled very efficiently by introducing auxiliary decision variables and constraints (see, e.g., Winston (2004)).

3.6. Combining criteria

Several other treatment plan evaluation criteria that have been proposed in the literature can be viewed as linear combinations of the elementary criteria that we have discussed in section 3.3. When we compare multi-criteria models for FMO based on linear combinations of elementary convex criteria, it is easy to see that these models can be obtained from multi-criteria models based on the individual elementary criteria by restricting the range of values that is considered for the criterion weights $\vec{\mu}$. Models based on such combined criteria thus impose an *a priori* trade-off between the corresponding elementary criteria, thereby significantly reducing the set of efficient treatment plans. In the absence of a fundamental basis for the quantification of the trade-off between the elementary criteria, it is preferable to consider the elementary criteria explicitly rather than the combined criteria. In the remainder of this section, we will address this issue more explicitly by studying combined criteria that have been proposed in the literature.

3.6.1. Alternative measure of EUD. For critical structures, Dale and Olsen (1997) suggested the use of a linear combination of the maximum and the mean dose as an evaluation criterion of a dose distribution:

$$\begin{aligned} \mathrm{EUD}^{\mathrm{D\&O}}(\vec{d};\alpha) &= \alpha \cdot \mathrm{gEUD}(\vec{d};\infty) + (1-\alpha)\cdot \mathrm{gEUD}(\vec{d};1) \\ &= \alpha \cdot \max_{j=1,\ldots,v} d_j + (1-\alpha)\cdot \frac{1}{v}\sum_{j=1}^{v} d_j \end{aligned}$$

where $\alpha \in [0, 1]$ depends on the critical structure. Thieke *et al* (2002) have further developed this approach and estimated values of the parameter α corresponding to several critical structures based on experimental data. However, due to the difficulties in accurately estimating the parameter α and the lack of an underlying theoretical motivation for using $\mathrm{EUD}^{\mathrm{D\&O}}$, it will be preferable to simply incorporate the maximum dose $\mathrm{gEUD}(\vec{d};\infty)$ and the mean dose $\mathrm{gEUD}(\vec{d};1)$ explicitly in a multi-criteria model for FMO without imposing an *a priori* trade-off between them.

3.6.2. Scaling EUD. Bortfeld *et al* (2003) propose to measure EUD criteria relative to predefined values. In particular, they propose to use

$$\frac{\underline{d} - \mathrm{gEUD}(\vec{d};-\infty)}{\underline{d}} \tag{9}$$

for targets, and

$$\frac{\mathrm{EUD}^{\mathrm{D\&O}}(\vec{d};\alpha) - \overline{d}}{\overline{d}} \tag{10}$$

for critical structures, where $\underline{d}$ and $\overline{d}$ are some pre-defined desirable levels of the criteria $\mathrm{gEUD}(\vec{d};-\infty)$ and $\mathrm{EUD}^{\mathrm{D\&O}}(\vec{d};\alpha)$, respectively. In a multi-criteria setting, this approach is equivalent to using the criteria $\mathrm{gEUD}(\vec{d};-\infty)$ and $\mathrm{EUD}^{\mathrm{D\&O}}(\vec{d};\alpha)$, and the discussion in section 3.6.1 applies.

3724

3.6.3. Probability of uncomplicated tumour control. The probability of uncomplicated tumour control, P_+, is a measure that has been proposed to make a trade-off between TCP values for targets and NTCP values for critical structures. In particular, P_+ is defined as

$$P_+(\vec{d}_1, \ldots, \vec{d}_S) = \prod_{s=1}^{T} \mathrm{TCP}_s(\vec{d}_s) \cdot \prod_{s=T+1}^{S} (1 - \mathrm{NTCP}_s(\vec{d}_s))$$

(see, e.g., Brahme 2001). Since larger values of P_+ are preferred to smaller values, in the remainder of this section we will consider its negative $-P_+$. Now note that we can decompose $-P_+$ as

$$-P_+(\vec{d}_1, \ldots, \vec{d}_S) = h\left(-\sum_{s=1}^{T} \ln \mathrm{TCP}_s(\vec{d}_s) - \sum_{s=T+1}^{S} \ln(1 - \mathrm{NTCP}_s(\vec{d}_s))\right)$$

where $h(z) = -\mathrm{e}^{-z}$. Theorem 2.2 says that the efficient frontier for a multi-criteria formulation of the FMO problem does not change if we replace the criterion $-P_+$ by the combined criterion

$$-\sum_{s=1}^{T} \ln \mathrm{TCP}_s(\vec{d}_s) - \sum_{s=T+1}^{S} \ln(1 - \mathrm{NTCP}_s(\vec{d}_s)).$$

As we have mentioned above, $-\ln \mathrm{TCP}_s(\vec{d}_s)$ is convex. Moreover, note that

$$\begin{aligned} -\ln\left(1 - \mathrm{NTCP}_s^{\mathrm{LS}}(\vec{d}_s; a, m, D_{50})\right) &= -\ln\left(1 - \Phi\left(\frac{\mathrm{gEUD}_s(\vec{d}_s; a) - D_{50}}{m D_{50}}\right)\right) \\ &= \breve{W}(\vec{d}_s; a, m, D_{50}) \end{aligned}$$

and

$$\begin{aligned} -\ln\left(1 - \mathrm{NTCP}_s^{\mathrm{A\&N}}(\vec{d}_s; a, \Delta)\right) &= \left(\frac{\mathrm{gEUD}_s(\vec{d}_s; a)}{\Delta}\right)^a \\ &= \frac{1}{\Delta^a} \cdot F_s^{\mathrm{g}}(\vec{d}_s; a) \end{aligned}$$

which means that $-\ln(1 - \mathrm{NTCP}_s(\vec{d}_s))$ is convex for both the LS and the A&N models. This then implies that we can explore a vastly larger set of efficient treatment plans by considering the individual criteria $-\ln \mathrm{TCP}_s(\vec{d}_s)$ and $-\ln(1 - \mathrm{NTCP}_s(\vec{d}_s))$. By table 1, the former criteria are in fact equivalent to $\mathrm{TCP}_s(\vec{d}_s)$. Similarly, by choosing $h(z) = -\ln(1 - z)$, the latter criteria are equivalent to $\mathrm{NTCP}_s(\vec{d}_s)$. Therefore, the treatment plans that are Pareto efficient with respect to the criterion P_+ can be found by considering the more elementary TCP and NTCP measures. However, the latter measures are preferable, since they do not impose an *a priori* trade-off between these measures. Finally, we note that any multi-criteria model based on P_+, TCP and NTCP measures can be transformed into an equivalent model based on convex separable criteria, leading to voxel-based model formulations for FMO.

As a last remark, the analysis of P_+ reveals a close relationship between NTCP and $\breve{W}$. In fact, if we consider measuring TCP according to the L&S model and define

$$\mathrm{TCP}^{\mathrm{L\&S}}(\vec{d}; a, m, D_{50}) = \Phi\left(\frac{\mathrm{gEUD}(\vec{d}; a) - D_{50}}{m D_{50}}\right)$$

3725

for $-\infty \leqslant a \leqslant 0$ and $m,\ D_{50} > 0$, we obtain that

$$\begin{aligned} -\ln\left(\mathrm{TCP}_s^{\mathrm{LS}}(\vec{d}_s; a, m, D_{50})\right) &= -\ln\left(\Phi\left(\frac{\mathrm{gEUD}_s(\vec{d}_s; a) - D_{50}}{m D_{50}}\right)\right) \\ &= -\ln\left(1 - \Phi\left(\frac{D_{50} - \mathrm{gEUD}_s(\vec{d}_s; a)}{m D_{50}}\right)\right) \\ &= \breve{W}(\vec{d}_s; a, m, D_{50}). \end{aligned}$$

The corresponding measure of P_+ is then given by

$$P_+^{\mathrm{L\&S}} = \prod_{s=1}^{T} \mathrm{TCP}_s^{\mathrm{L\&S}}(\vec{d}_s) \cdot \prod_{s=T+1}^{S} \left(1 - \mathrm{NTCP}_s^{\mathrm{L\&S}}(\vec{d}_s)\right) = \prod_{s=1}^{S} \breve{w}(\vec{d}_s)$$

establishing a very close relationship between P_+ and the alternative to the criterion by Wu *et al* (2002).

4. Discussion and conclusions

We have a presented a unifying framework for studying multi-criteria models for FMO. Casting FMO models as multi-criteria models has allowed us to establish the equivalence between many sets of treatment plan evaluation criteria that have been proposed in the literature. We have also established conditions under which nonconvex criteria can be transformed into convex criteria without changing the set of Pareto efficient treatment plans.

The framework identifies which of the proposed treatment plan evaluation criteria are truly different. This enables researchers to focus their efforts on comparing non-equivalent models. We have shown that using the 'biological' criteria considered in this paper, which include EUD, gEUD, TCP, NTCP, P_+, as well as two sigmoidal transformations of EUD, only two distinct Pareto efficient frontiers exist. We have shown that these frontiers can be obtained using equivalent 'physical' convex voxel-based criteria. This result should alleviate concerns on the pros and cons of using biological versus physical criteria that was recently debated in the literature, as they are in essence equivalent (see, e.g., Wu *et al* (2002), Amols and Ling 2002, Thieke *et al* (2003c)). Clearly, many criteria still remain to be analysed. To avoid duplication of efforts, in the pursuit of new treatment plan evaluation criteria one should consider the unifying framework to determine whether potentially new criteria indeed yield different Pareto frontiers and therefore truly new models.

There exist treatment plan evaluation criteria that are fundamentally nonconvex, i.e., transformations to convex criteria using the framework do not exist. Examples of these are traditional dose-volume histogram criteria (see, e.g., Deasy (1997)) and certain measures of delivery efficiency such as the number of beam orientations used (see, e.g., Bortfeld and Schlegel (1993)). The consequence of including nonconvex treatment plan evaluation criteria into a multi-criteria optimization framework is that the Pareto efficient frontier may be nonconvex or even disconnected. This makes the comparison of multi-criteria models based on these criteria with multi-criteria models based on biological or voxel-based physical criteria impractical given the current state-of-the-art in solving the associated optimization problems (see, e.g., Lee *et al* (2003)). In particular, when using heuristic approaches for solving problems that are aimed at identifying points on the Pareto efficient frontier, it cannot be guaranteed that the obtained solutions are indeed Pareto efficient (see, e.g., Schreibmann *et al* (2004)). With respect to dose-volume histograms, Romeijn *et al* (2003) have recently proposed an alternative

family of convex criteria that measure the tail mean of a differential dose-volume histogram. Although this means that these criteria could in principle be incorporated into convex multi-criteria optimization models, no equivalent voxel-based criteria exist. However, given the widespread interest in and potential value of traditional dose-volume histogram criteria and beam number and orientation, future research will address nonconvex multi-criteria models for FMO.

Appendix A

The following theorem shows that Niemierko's original EUD (see Niemierko (1997)) is a concave function of the dose distribution.

Theorem A.1. *The function* $\mathrm{EUD}(\vec{d};\alpha)$ *is concave in* $\vec{d}$ *when* $\alpha > 0$.

Proof. For notational simplicity, let us assume that the unit of dose is normalized to make $\alpha = 1$ and define

$$G(\vec{d}) = \mathrm{EUD}(\vec{d};1) = -\ln\left(\frac{1}{v}\sum_{j=1}^{v} \mathrm{e}^{-d_j}\right).$$

The first derivatives of this function are equal to

$$\frac{\partial G(\vec{d})}{\partial d_k} = \frac{\mathrm{e}^{-d_k}}{\sum_{j=1}^{v} \mathrm{e}^{-d_j}} \qquad k = 1,\ldots,v$$

which implies that G is nondecreasing in d_k for all $k = 1,\ldots,v$. The second derivatives of G are given by

$$\frac{\partial^2 G(\vec{d})}{\partial d_k^2} = \frac{\mathrm{e}^{-2d_k} - \mathrm{e}^{-d_k}\left(\sum_{j=1}^{v} \mathrm{e}^{-d_j}\right)}{\left(\sum_{j=1}^{v} \mathrm{e}^{-d_j}\right)^2} \qquad k = 1,\ldots,v$$

$$\frac{\partial^2 G(\vec{d})}{\partial d_k \partial d_{k'}} = \frac{\mathrm{e}^{-d_k - d_{k'}}}{\left(\sum_{j=1}^{v} \mathrm{e}^{-d_j}\right)^2} \qquad k,k' = 1,\ldots,v;\ k \neq k'.$$

Ignoring the common, positive denominator in the partial derivatives and denoting the resulting matrix by $\hat{H}(\vec{d})$, we can study the definiteness of this matrix by studying the sign of the quadratic form $y^{\mathrm{T}}\hat{H}(\vec{d})y$ for all vectors y:

$$\begin{aligned}
y^{\mathrm{T}}\hat{H}(\vec{d})y &= \left(\sum_{k=1}^{v} \mathrm{e}^{-d_k} y_k\right)\cdot\left(\sum_{k'=1}^{v} \mathrm{e}^{-d_{k'}} y_{k'}\right) - \left(\sum_{k=1}^{v} \mathrm{e}^{-d_k}\right)\cdot\left(\sum_{k'=1}^{v} \mathrm{e}^{-d_{k'}} y_{k'}^2\right) \\
&= \sum_{k=1}^{v}\sum_{k'=1}^{v} \mathrm{e}^{-d_k - d_{k'}} y_k y_{k'} - \sum_{k=1}^{v}\sum_{k'=1}^{v} \mathrm{e}^{-d_k - d_{k'}} y_{k'}^2 \\
&= 2\sum_{k=1}^{v}\sum_{k'=k+1}^{v} \mathrm{e}^{-d_k - d_{k'}} y_k y_{k'} - \sum_{k=1}^{v}\sum_{k'=1}^{k-1} \mathrm{e}^{-d_k - d_{k'}} y_{k'}^2 - \sum_{k=1}^{v}\sum_{k'=k+1}^{v} \mathrm{e}^{-d_k - d_{k'}} y_{k'}^2 \\
&= 2\sum_{k=1}^{v}\sum_{k'=k+1}^{v} \mathrm{e}^{-d_k - d_{k'}} y_k y_{k'} - \sum_{k=1}^{v}\sum_{k'=k+1}^{v} \mathrm{e}^{-d_{k'} - d_k} y_k^2 - \sum_{k=1}^{v}\sum_{k'=k+1}^{v} \mathrm{e}^{-d_k - d_{k'}} y_{k'}^2
\end{aligned}$$

3727

$$
\begin{aligned}
&= 2\sum_{k=1}^{v}\sum_{k'=k+1}^{v}\left(e^{-(d_k+d_{k'})/2}y_k\right)\cdot\left(e^{-(d_k+d_{k'})/2}y_{k'}\right)\\
&\quad -\sum_{k=1}^{v}\sum_{k'=k+1}^{v}\left(e^{-(d_k+d_{k'})/2}y_k\right)^2-\sum_{k=1}^{v}\sum_{k'=k+1}^{v}\left(e^{-(d_k+d_{k'})/2}y_{k'}\right)^2\\
&= -2\sum_{k=1}^{v}\sum_{k'=k+1}^{v}\left(\left(e^{-(d_k+d_{k'})/2}y_k\right)-\left(e^{-(d_k+d_{k'})/2}y_{k'}\right)\right)^2\\
&\leqslant 0
\end{aligned}
$$

for all y, which implies that G is concave. □

Appendix B

The following lemma is useful to derive properties of the criterion function W proposed by Wu *et al* (2002).

Lemma B.1. *Define the function*

$$\xi_k(z) = \ln(1+z^k) \qquad z \geqslant 0$$

for $k \in \mathbb{R}\backslash\{0\}$. *If* $k < 0$, ξ_k *is convex on* $\mathbb{R}_+$. *If* $0 < k \leqslant 1$, ξ_k *is concave on* $\mathbb{R}_+$. *If* $k > 1$, ξ_k *is convex on* $[0, \sqrt[k]{k-1}]$ *and concave on* $[\sqrt[k]{k-1}, \infty)$.

Proof. The first derivative of ξ_k is equal to

$$\xi_k'(z) = \frac{kz^{k-1}}{1+z^k}$$

which means that ξ_k is increasing if $k > 0$ and decreasing if $k < 0$. Its second derivative is equal to

$$\xi_k''(z) = \frac{kz^{k-2}(k-1-z^k)}{(1+z^k)^2}.$$

If $k < 0$, we have that ξ_k is convex if

$$z^k \geqslant k-1$$

which means that ξ_k is convex on $\mathbb{R}_+$ whenever $k < 0$. If $k > 0$, we have that ξ_k is convex if

$$z^k \leqslant k-1$$

and concave otherwise. This means that ξ_k is concave on $\mathbb{R}_+$ if $0 < k \leqslant 1$. If $k > 1$, ξ_k is convex on $[0, \sqrt[k]{k-1}]$ and concave on $[\sqrt[k]{k-1}, \infty)$. □

This then leads to the following

Theorem B.2. *The function* $W(\vec{d}; a, k, \text{gEUD}^0)$ *is concave in* $\vec{d}$ *whenever* $-\infty \leqslant a \leqslant 0$ *and* $k, \text{gEUD}^0 > 0$. *The function* $W(\vec{d}; a, k, \text{gEUD}^0)$ *is convex in* $\vec{d}$ *in the region where* $\text{gEUD}(\vec{d}; a) \leqslant \text{gEUD}_0\sqrt[k]{k-1}$ *whenever* $1 \leqslant a \leqslant \infty, k > 1$, *and* $\text{gEUD}^0 > 0$.

Proof. If $-\infty \leqslant a \leqslant 0$, we have that

$$W(\vec{d}; a, k, \text{gEUD}_0) = \xi_{-k}\left(\frac{\text{gEUD}(\vec{d}; a)}{\text{gEUD}_0}\right).$$

Since $k > 0$, lemma B.1 says that ξ_{-k} is a decreasing and convex function. Since $a \leqslant 0$, $\mathrm{gEUD}(\vec{d}; a)$ is a concave function of $\vec{d}$, and we conclude that W is a convex function of $\vec{d}$.

If $1 \leqslant a \leqslant \infty$, we have that

$$W(\vec{d}; a, k, \mathrm{gEUD}_0) = \xi_k\left(\frac{\mathrm{gEUD}(\vec{d}; a)}{\mathrm{gEUD}_0}\right).$$

If $k > 1$, lemma B.1 says that ξ_k is an increasing function that is convex on $[0, \sqrt[k]{k-1}]$. Moreover, since $a \geqslant 1$, $\mathrm{gEUD}(\vec{d}; a)$ is a convex function of $\vec{d}$. Therefore, we conclude that W is a convex function of $\vec{d}$ in the region where $\mathrm{gEUD}(\vec{d}; a) \leqslant \mathrm{gEUD}_0\sqrt[k]{k-1}$. □

In the last theorem of this appendix, we derive properties of the alternative criterion $\breve{W}$.

Theorem B.3. *The function* $\breve{W}(\vec{d}; a, \sigma, \mathrm{gEUD}^0)$ *is convex in* $\vec{d}$ *whenever* $-\infty \leqslant a \leqslant 0$ *or* $1 \leqslant a \leqslant \infty$ *and* $\sigma, \mathrm{gEUD}^0 > 0$.

Proof. Defining

$$\psi(z) = -\ln(1 - \Phi(z))$$

we can write

$$\breve{W}(\vec{d}; a, \sigma, \mathrm{gEUD}^0) = \begin{cases} \psi\left(\dfrac{\mathrm{gEUD}^0 - \mathrm{gEUD}(\vec{d}; a)}{\sigma \cdot \mathrm{gEUD}^0}\right) & \text{if } -\infty \leqslant a \leqslant 0 \\ \psi\left(\dfrac{\mathrm{gEUD}(\vec{d}; a) - \mathrm{gEUD}^0}{\sigma \cdot \mathrm{gEUD}^0}\right) & \text{if } 1 \leqslant a \leqslant \infty. \end{cases}$$

It is straightforward to derive that

$$\psi'(z) = \frac{\Phi'(z)}{1 - \Phi(z)}$$

where $\Phi'(z)$ is the probability density function of a standard normal random variable. Clearly, this implies that ψ is increasing. Moreover, $\psi'(z)$ is the so-called failure rate of the standard normal distribution, an applied probability concept that is often used in the area of reliability. It is known that this failure rate is increasing (see, e.g., Park (1987)), which implies that ψ is convex. Now recall that $-\mathrm{gEUD}(a)$ is convex for $-\infty \leqslant a \leqslant 0$ and $\mathrm{gEUD}(a)$ is convex for $1 \leqslant a \leqslant \infty$; we obtain that $\breve{W}$ is an increasing convex function of a convex function and therefore convex. □

References

Abramowitz M and Stegun I A (ed) 1965 *Handbook of Mathematical Functions* (New York: Dover)

Alber M and Nüsslin F 1999 An objective function for radiation treatment optimization based on local biological measures *Phys. Med. Biol.* **44** 479–93

Alber M and Nüsslin F 2001 A representation of an NTCP function for local complication mechanisms *Phys. Med. Biol.* **46** 439–47

Amols H I and Ling C C 2002 EUD but not QED *Int. J. Radiat. Oncol. Biol. Phys.* **52** 1–2

Bortfeld T, Küfer K-H, Monz M, Scherrer A, Thieke C and Trinkhaus H 2003 Intensity-modulated radiotherapy: a large scale multi-criteria programming problem *Berichte des Fraunhofer ITWM* 43 (Kaiserslautern, Germany: Fraunhofer Insitut Techno- und Wirtschaftsmathematik)

Bortfeld T and Schlegel W 1993 Optimization of beam orientations in radiation therapy: some theoretical considerations *Phys. Med. Biol.* **38** 291–304

Brahme A 2001 Individualizing cancer treatment: biological optimization models in treatment planning and delivery *Int. J. Radiat. Oncol. Biol. Phys.* **49** 327–37

3729

Brahme A and Agren A K 1987 Optimal dose distribution for eradication of heterogeneous tumours *Acta Oncol.* **26** 377–85

Choi B and Deasy J O 2002 The generalized equivalent uniform dose function as a basis for intensity-modulated treatment planning *Phys. Med. Biol.* **47** 3579–89

Dale E and Olsen D R 1997 Specification of the dose to organs at risk in external beam radiotherapy *Acta Oncol.* **36** 129–35

Deasy J O 1997 Multiple local minima in radiotherapy optimization problem with dose-volume constraints *Med. Phys.* **24** 1157–61

Goitein M 1987 Tumour control probability for an inhomogeneously irradiated target volume *Evaluation of Treatment Planning for Particle Beam Radiotherapy* (Bethesda, MD: National Cancer Institute)

Hamacher H W and Küfer K-H 2002 Inverse radiation therapy planning—a multiple objective optimization approach *Discrete Appl. Math.* **118** 145–61

Küfer K-H, Hamacher H W and Bortfeld T R 2000 A multicriteria optimization approach for inverse radiotherapy planning *Proc. XIIIth ICCR* ed T R Bortfeld and W Schlegel *(Heidelberg, Germany)* pp 26–9

Kutcher G J and Burman C 1989 Calculation of complication probability factors for non-uniform normal tissue irradiation: the effective volume method *Int. J. Radiat. Oncol. Biol. Phys.* **16** 1623–30

Lahanas M, Schreibmann E and Baltas D 2003 Multiobjective inverse planning for intensity modulated radiotherapy with constraint-free gradient-based optimization algorithms *Phys. Med. Biol.* **48** 2843–71

Lee E K, Fox T and Crocker I 2003 Integer programming applied to intensity-modulated radiation treatment planning *Ann. Oper. Res.* **119** 165–81

Lyman J T 1985 Complication probability as assessed from dose-volume histograms *Radiat. Res.* **8** S13–9

Lyman J T and Wolbarst A B 1987a Optimization of radiation therapy III: a method of assessing complication probabilities from dose-volume histograms *Int. J. Radiat. Oncol. Biol. Phys.* **13** 103–9

Lyman J T and Wolbarst A B 1987b Optimization of radiation therapy IV: a dose-volume histogram reduction algorithm *Int. J. Radiat. Oncol. Biol. Phys.* **13** 103–9

McGary J E, Grant III W and Woo S Y 2000 Applying the equivalent uniform dose formulation based on the linear-quadratic model to inhomogeneous tumor dose distribution: caution for analyzing and reporting *J. Appl. Clin. Med. Phys.* **1** 126–37

Miettinen K 1999 *Nonlinear Multiobjective Optimization* (Boston, MA: Kluwer)

Niemierko A 1997 Reporting and analyzing dose distributions: a concept of equivalent uniform dose *Med. Phys.* **24** 103–10

Niemierko A 1999 A generalized concept of equivalent uniform dose (EUD) *AAPM Annual Meeting, (Nashville, Tennessee)* Abstract *Med. Phys.* **26** 1100

Niemierko A and Goitein M 1991 Calculation of normal tissue complication probability and dose-volume histogram reduction schemes for tissues with a critical element architecture *Radiother. Oncol.* **20** 166–76

Pareto V 1971 *Manual of Political Economy* (New York: Kelley) (Translation of Manuale di economia politica, 1906)

Park K S 1987 *Human Reliability* (Amsterdam: Elsevier)

Romeijn H E, Ahuja R K, Dempsey J F, Kumar A and Li J G 2003 A novel linear programming approach to fluence map optimization for intensity modulated radiation therapy treatment planning *Phys. Med. Biol.* **48** 3521–42

Schreibmann E, Lahanas M, Xing L and Baltas D 2004 Multiobjective evolutionary optimization of the number of beams, their orientations and weights for intensity-modulated radiation therapy *Phys. Med. Biol.* **49** 747–70

Schultheiss T E, Orton C G and Peck R A 1983 Models in radiotherapy: volume effects *Med. Phys.* **10** 410–5

Shepard D M, Ferris M C, Olivera G H and Mackie T R 1999 Optimizing the delivery of radiation therapy to cancer patients *SIAM Rev.* **41** 721–44

Stavrev P, Hristov D, Warkentin B, Sham E, Stavreva N and Fallone B G 2003 Inverse treatment planning by physically constrained minimization of a biological objective function *Med. Phys.* **30** 2948–58

Stavreva N A, Stavrev P, Warkentin B and Fallone B G 2003 Investigating the effect of cell repopulation on the tumor response to fractionated external radiotherapy *Med. Phys.* **30** 735–42

Steuer R 1986 *Multicriteria Optimization: Theory, Computation and Applications* (New York: Wiley)

Thieke C, Bortfeld T and Küfer K-H 2002 Characterization of dose distributions through the Max and Mean dose concept *Acta Oncol.* **41** 158–61

Thieke C, Bortfeld T and Küfer K H 2003a New optimization concepts in inverse treatment planning *Pre-meeting Workshop on 'Optimization of IMRT', 7th Biennial ESTRO Meeting on Physics and Radiation Technology for Clinical Radiotherapy (Geneva, Switzerland, Sept. 2003)* Abstract *Radiother. Oncol.* **68** (Suppl. 1)S2

Thieke C, Bortfeld T, Niemierko A, Küfer K H and Nill S 2003b Multicriteria optimization in inverse radiotherapy planning *7th Biennial ESTRO Meeting on Physics and Radiation Technology for Clinical Radiotherapy (Geneva, Switzerland, Sept. 2003)* Abstract *Radiother. Oncol.* **68** (Suppl. 1)S44

3730

Thieke C, Bortfeld T, Niemierko A and Nill S 2003c From physical dose constraints to equivalent uniform dose constraints in inverse radiotherapy planning *Med. Phys.* **30** 2332–9

Tsien C, Eisbruch A, McShan D, Kessler M, Marsh R and Fraass B 2003 Intensity-modulated radiation therapy (IMRT) for locally advanced paranasal sinus tumors: incorporating clinical decisions in the optimization process *Int. J. Radiat. Oncol. Biol. Phys.* **55** 776–84

Winston W L 2004 *Operations Research: Applications and Algorithms* 4th edn (Belmont, CA: Thomson Learning Inc.)

Withers H R and McBride W H 1998 Biologic basis of radiation therapy Principles and Practice of Radiotherapy ed C A Perez and L W Brady (Philadelphia, PA: Williams and Wilkins) pp 79–118 chapter 2

Wolbarst A B 1984 Optimization of radiation therapy II: the critical-voxel method *Int. J. Radiat. Oncol. Biol. Phys.* **10** 741–5

Wu Q, Mohan R, Niemierko A and Schmidt-Ullrich R 2002 Optimization of intensity-modulated radiotherapy plans based on the equivalent uniform dose *Int. J. Radiat. Oncol. Biol. Phys.* **52** 224–35

Zaider M and Minerbo G N 2000 Tumour control probability: a formulation applicable to any temporal protocol of dose delivery *Phys. Med. Biol.* **45** 279–93

INSTITUTE OF PHYSICS PUBLISHING PHYSICS IN MEDICINE AND BIOLOGY

Phys. Med. Biol. 49 (2004) N7–N19 PII: S0031-9155(04)65857-X

NOTE

Optimal leaf sequencing with elimination of tongue-and-groove underdosage

Srijit Kamath[1], Sartaj Sahni[1], Jatinder Palta[2], Sanjay Ranka[1] and Jonathan Li[2]

[1] Department of Computer and Information Science and Engineering, University of Florida, Gainesville, FL, USA
[2] Department of Radiation Oncology, University of Florida, Gainesville, FL, USA

E-mail: srkamath@cise.ufl.edu

Received 8 July 2003
Published 16 January 2004
Online at stacks.iop.org/PMB/49/N7 (DOI: 10.1088/0031-9155/49/3/N01)

Abstract

The individual leaves of a multileaf collimator (MLC) have a tongue-and-groove or stepped-edge design to minimize leakage radiation between adjacent leaves. This design element has a drawback in that it creates areas of underdosages in intensity-modulated photon beams unless a leaf trajectory is specifically designed such that for any two adjacent leaf pairs, the direct exposure under the tongue-and-groove is equal to the lower of the direct exposures of the leaf pairs. In this work, we present a systematic study of the optimization of a leaf sequencing algorithm for segmental multileaf collimator beam delivery that completely eliminates areas of underdosages due to tongue-and-groove or stepped-edge design of the MLC. Simultaneous elimination of tongue-and-groove effect and leaf interdigitation is also studied. This is an extension of our previous work (Kamath *et al* 2003a *Phys. Med. Biol.* **48** 307) in which we described a leaf sequencing algorithm that is optimal for monitor unit (MU) efficiency under most common leaf movement constraints that include minimum leaf separation. Compared to our previously published algorithm (without constraints), the new algorithms increase the number of sub-fields by approximately 21% and 25%, respectively, but are optimal in MU efficiency for unidirectional schedules.

1. Introduction

Intensity-modulated radiation therapy (IMRT) delivered with a multileaf collimator (MLC) in the step-and-shoot mode uses multiple static MLC segments to achieve intensity modulation. The sides of each leaf of a MLC have a protruding tongue or a step on one side that fits into a similar groove of the adjacent leaf. This results in different radiological path lengths across different parts of the leaves. Galvin *et al* (1993a) first described that the different

0031-9155/04/030007+13$30.00

3752

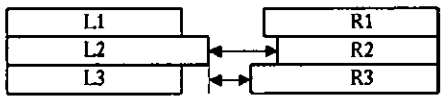


Figure 1. Inter-pair minimum separation constraint.

radiological path lengths manifest themselves as varying doses in a plane perpendicular to the leaf motion. The low dose region between two adjacent leaves was classified as the tongue-and-groove effect. In an IMRT treatment using a MLC, the tongue-and-groove effect occurs when the tongue, or the groove or both for the most time during treatment delivery cover the overlapping region between two adjacent pairs of leaves. As pointed out by many investigators, the tongue-and-groove arrangement always results in underdosages of as much as 10–25% in the treatment fields in both segmented multileaf collimation (SMLC) and dynamic multileaf collimation (DMLC) (Galvin *et al* 1993a, 1993b, Chui *et al* 1994, Mohan 1995, Wang *et al* 1996, Sykes and Williams 1998).

Several recent publications (van Santvoort and Heijmen 1996, Webb *et al* 1997, Convery and Webb 1998, Dirkx *et al* 1998, Xia and Verhey 1998) have shown that the tongue-and-groove effect can be significantly reduced by synchronization of the leaves. However, the cost of leaf synchronization is usually an increase in the total number of sub fields and monitor units. van Santvoort and Heijmen (1996) propose an algorithm to eliminate tongue-and-groove effects for DMLC treatment plans. Although they note that their algorithm increases the number of monitor units, they do not examine the optimality or suboptimality of the plans they obtain. We recently published a paper (Kamath *et al* 2003a) that gave mathematical formalisms and rigorous proofs of leaf sequencing algorithms for segmental multileaf collimation, which maximize MU efficiency. We proved that our leaf sequencing algorithms that explicitly account for minimum leaf separation obtain feasible unidirectional solutions that are optimal. We now extend that work to develop algorithms that explicitly account for leaf interdigitation and the tongue-and-groove effect and are optimal in MU efficiency for unidirectional schedules. The model used here is the same as that used in Kamath *et al* (2003a) and the reader is referred to this paper for fundamental definitions and algorithms. We show also that the algorithm of van Santvoort and Heijmen (1996) obtains optimal DMLC treatment schedules.

2. Optimal algorithm with interdigitation and tongue-and-groove constraints

2.1. Interdigitation constraint

In practical situations, there are some constraints on the movement of the leaves. The minimum separation constraint requires that opposing pairs of leaves be separated by at least some distance ($S_{\min}$) at all times during beam delivery. In some MLCs this constraint is applied not only to opposing pairs of leaves, but also to opposing leaves of neighbouring pairs. For example, in figure 1, $L1$ and $R1$, $L2$ and $R2$, $L3$ and $R3$, $L1$ and $R2$, $L2$ and $R1$, $L2$ and $R3$, $L3$ and $R2$ are pairwise subject to the constraint. We use the term *intra-pair minimum separation constraint* to refer to the constraint imposed on an opposing pair of leaves and *inter-pair minimum separation constraint* to refer to the constraint imposed on opposing leaves of neighbouring pairs. The *inter-pair minimum separation constraint* with $S_{\min} = 0$ is of special interest and is referred to as the *interdigitation constraint*.

3733

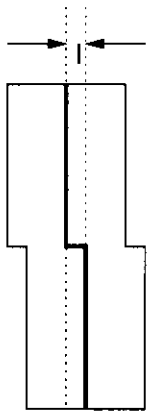


Figure 2. Cross section of leaves.

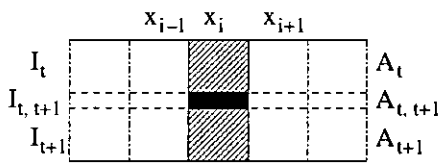


Figure 3. Tongue-and-groove effect.

2.2. Tongue-and-groove underdosage effect

In most commercially available MLCs, there is a tongue-and-groove arrangement at the interface between adjacent leaves. A cross section of two adjacent leaves is depicted in figure 2. The width of the tongue-and-groove region is l. The area under this region gets underdosed due to the mechanical arrangement. Figure 3 shows a beams-eye view of the region to be treated by two adjacent leaf pairs, t and $t+1$. Consider the shaded rectangular areas $A_t(x_i)$ and $A_{t+1}(x_i)$ that require exactly $I_t(x_i)$ and $I_{t+1}(x_i)$ MUs to be delivered, respectively. The tongue-and-groove overlap area between the two leaf pairs over the sample point x_i, $A_{t,t+1}(x_i)$, is coloured black. Let the amount of MUs delivered in $A_{t,t+1}(x_i)$ be $I_{t,t+1}(x_i)$. Ignoring leaf transmission, the following lemma is a consequence of the fact that $A_{t,t+1}(x_i)$ is exposed only when both $A_t(x_i)$ and $A_{t+1}(x_i)$ are exposed.

Lemma 1. $I_{t,t+1}(x_i) \leqslant \min\{I_t(x_i), I_{t+1}(x_i)\}, 0 \leqslant i \leqslant m, 1 \leqslant t < n$, *where m is the number of sample points along each row and n is the number of leaf pairs.*

Schedules in which $I_{t,t+1}(x_i) = \min\{I_t(x_i), I_{t+1}(x_i)\}$ are said to be free of tongue-and-groove underdosage effects.

Let $I_{tl}(x_i)$ and $I_{tr}(x_i)$ respectively denote the number of MUs delivered when the left and right leaves of pair t pass the point x_i during their left to right movement. Unless treatment schedules are carefully designed, it is possible that $I_{t,t+1}(x_i) \ll \min\{I_t(x_i), I_{t+1}(x_i)\}$ for some i and t. For example, in a schedule in which $I_{tr}(x_i) = 30$, $I_{tl}(x_i) = 50$, $I_{(t+1)r}(x_i) = 50$ and $I_{(t+1)l}(x_i) = 60$, we have $I_{t,t+1}(x_i) = I_{tl}(x_i) - I_{(t+1)r}(x_i) = 50 - 50 = 0$. Note that in this case, $\min\{I_t(x_i), I_{t+1}(x_i)\} = I_{(t+1)l}(x_i) - I_{tl}(x_i) = 60 - 50 = 10$. It is clear from this example that $I_{t,t+1}(x_i)$ could be 0 even when $\min\{I_t(x_i), I_{t+1}(x_i)\}$ is arbitrarily large.

3712

2.3. Algorithms

Kamath *et al* (2003a) present an algorithm that generates a schedule that satisfies inter-pair minimum separation constraint. The schedule is optimal in therapy time. However, it does not account for the tongue-and-groove effect. In this section, we present two algorithms. Algorithm TONGUEANDGROOVE generates minimum therapy time unidirectional schedules that are free of tongue-and-groove underdosage and maybe used for MLCs that do not have a interdigitation constraint. Algorithm TONGUEANDGROOVE-ID generates minimum therapy time unidirectional schedules that are free of tongue-and-groove underdosage while simultaneously satisfying the interdigitation constraint and is for MLCs that have an interdigitation constraint.

The following lemma provides a necessary and sufficient condition for a unidirectional schedule to be free of tongue-and-groove underdosage effects.

Lemma 2. *A unidirectional schedule is free of tongue-and-groove underdosage effects if and only if,*

(a) $I_t(x_i) = 0$ *or* $I_{t+1}(x_i) = 0$, *or*
(b) $I_{tr}(x_i) \leqslant I_{(t+1)r}(x_i) \leqslant I_{(t+1)l}(x_i) \leqslant I_{tl}(x_i)$, *or*
(c) $I_{(t+1)r}(x_i) \leqslant I_{tr}(x_i) \leqslant I_{tl}(x_i) \leqslant I_{(t+1)l}(x_i)$,

$0 \leqslant i \leqslant m$, $1 \leqslant t < n$.

Proof. It is easy to see that any schedule that satisfies the above conditions is free of tongue-and-groove underdosage effects. So what remains is for us to show that every schedule that is free of tongue-and-groove underdosage effects satisfies the above conditions. Consider any such schedule. If condition (a) is satisfied at every i and t, the proof is complete. So assume i and t such that $I_t(x_i) \neq 0$ and $I_{t+1}(x_i) \neq 0$ exist. We need to show that either (b) or (c) is true for this value of i and t. Since the schedule is free of tongue-and-groove effects,

$$I_{t,t+1}(x_i) = \min\{I_t(x_i), I_{t+1}(x_i)\} > 0. \tag{1}$$

From the unidirectional constraint, it follows that $A_{t,t+1}(x_i)$ first gets exposed when both right leaves pass x_i, and it remains exposed till the first of the left leaves passes x_i. Further, if a left leaf passes x_i before a neighbouring right leaf passes x_i, $A_{t,t+1}(x_i)$ is not exposed at all. So,

$$I_{t,t+1}(x_i) = \max\{0, I_{(t,t+1)l}(x_i) - I_{(t,t+1)r}(x_i)\} \tag{2}$$

where $I_{(t,t+1)r}(x_i) = \max\{I_{tr}(x_i), I_{(t+1)r}(x_i)\}$ and $I_{(t,t+1)l}(x_i) = \min\{I_{tl}(x_i), I_{(t+1)l}(x_i)\}$. From (1) and (2), it follows that

$$I_{t,t+1}(x_i) = I_{(t,t+1)l}(x_i) - I_{(t,t+1)r}(x_i). \tag{3}$$

Consider the case $I_t(x_i) \geqslant I_{t+1}(x_i)$. Suppose that $I_{tr}(x_i) > I_{(t+1)r}(x_i)$. It follows that $I_{(t,t+1)r}(x_i) = I_{tr}(x_i)$ and $I_{(t,t+1)l}(x_i) = I_{(t+1)l}(x_i)$. Now from (3), we get

$$\begin{aligned} I_{t,t+1}(x_i) &= I_{(t+1)l}(x_i) - I_{tr}(x_i) \\ &< I_{(t+1)l}(x_i) - I_{(t+1)r}(x_i) \\ &= I_{t+1}(x_i) \\ &\leqslant I_t(x_i). \end{aligned} \tag{4}$$

So $I_{t,t+1}(x_i) < \min\{I_t(x_i), I_{t+1}(x_i)\}$, which contradicts (1). So

$$I_{tr}(x_i) \leqslant I_{(t+1)r}(x_i). \tag{5}$$

 3735

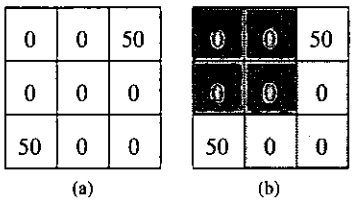


Figure 4. The intensity matrix shown in (a) can be treated using a single segment with 50 MUs as shown in (b). Areas shaded dark are covered by left leaves and those shaded light are covered by right leaves. Areas not shaded are exposed. Interdigitation constraint violation occurs though there is no tongue-and-groove violation.

Now, suppose that $I_{tl}(x_i) < I_{(t+1)l}(x_i)$. From $I_t(x_i) \geqslant I_{t+1}(x_i)$, it follows that $I_{(t,t+1)l}(x_i) = I_{tl}(x_i)$ and $I_{(t,t+1)r}(x_i) = I_{(t+1)r}(x_i)$. Hence, from (3), we get

$$\begin{aligned} I_{t,t+1}(x_i) &= I_{tl}(x_i) - I_{(t+1)r}(x_i) \\ &< I_{(t+1)l}(x_i) - I_{(t+1)r}(x_i) \\ &= I_{t+1}(x_i) \\ &\leqslant I_t(x_i). \end{aligned} \tag{6}$$

So $I_{t,t+1}(x_i) < \min\{I_t(x_i), I_{t+1}(x_i)\}$, which contradicts (1). So

$$I_{tl}(x_i) \geqslant I_{(t+1)l}(x_i). \tag{7}$$

From (5) and (7), we can conclude that when $I_t(x_i) \geqslant I_{t+1}(x_i)$, (b) is true. Similarly one can show that when $I_{t+1}(x_i) \geqslant I_t(x_i)$, (c) is true. □

Lemma 2 is equivalent to saying that the time period for which a pair of leaves (say pair t) exposes the region $A_{t,t+1}(x_i)$ is completely contained by the time period for which pair $t+1$ exposes region $A_{t,t+1}(x_i)$, or vice versa, whenever $I_t(x_i) \neq 0$ and $I_{t+1}(x_i) \neq 0$. Note that if either $I_t(x_i)$ or $I_{t+1}(x_i)$ is zero the containment is not necessary. We will refer to the necessary and sufficient condition of lemma 2 as the *tongue-and-groove constraint condition*. Schedules that satisfy this condition will be said to satisfy the tongue-and-groove constraint. van Santvoort and Heijmen (1996) present an algorithm that generates schedules that satisfy the tongue-and-groove constraint for DMLC.

Xia and Verhey (1998) claim that every schedule that violates the interdigitation constraint also violates the tongue-and-groove constraint. We demonstrate with a counterexample that this is not necessarily the case. The intensity matrix shown in figure 4(a) can be exposed in a single segment as shown in figure 4(b). The segment is free of tongue-and-groove constraint violations, while it clearly violates the interdigitation constraint.

2.3.1. Elimination of tongue-and-groove effect. The schedule generated by Algorithm MULTIPAIR (Kamath *et al* 2003a) may violate the tongue-and-groove constraint. If the schedule has no tongue-and-groove constraint violations, it is the desired optimal schedule. If there are violations in the schedule, we eliminate all violations of the tongue-and-groove constraint starting from the left end, i.e., from x_0. To eliminate the violations, we modify those plans of the schedule that cause the violations. We scan the schedule from x_0 along the positive x direction looking for the least x_w at which there exist leaf pairs $u, t, t \in \{u-1, u+1\}$, that violate the constraint at x_w. After rectifying the violation at x_w we look for other violations. Since the process of eliminating a violation at x_w, may at times, lead to new violations at

3736

Algorithm TONGUEANDGROOVE

(i) $x = x_0$

(ii) While (there is a tongue-and-groove violation) do

(iii) Find the least x_w, $x_w \geq x$, such that there exist leaf pairs u, $u+1$, that violate the tongue-and-groove constraint at x_w.

(iv) Modify the schedule to eliminate the violation between leaf pairs u and $u+1$.

(v) $x = x_w$

(vi) End While

Figure 5. Obtaining a schedule under the tongue-and-groove constraint.

x_w, we need to search afresh from x_w every time a modification is made to the schedule. However, we will prove a bound of $O(n)$ on the number of violations that can occur at x_w. After eliminating all violations at a particular sample point, x_w, we move to the next point, i.e., we increment w and look for possible violations at the new point. We continue the scanning and modification process until no tongue-and-groove constraint violations exist. Algorithm TONGUEANDGROOVE (figure 5) outlines the procedure.

Let $M = ((I_{1l}, I_{1r}), (I_{2l}, I_{2r}), \ldots, (I_{nl}, I_{nr}))$ be the schedule generated by algorithm MULTIPAIR for the desired intensity profile.

Let $N(p) = ((I_{1lp}, I_{1rp}), (I_{2lp}, I_{2rp}), \ldots, (I_{nlp}, I_{nrp}))$ be the schedule obtained after step (iv) of algorithm TONGUEANDGROOVE is applied p times to the input schedule M. Note that $M = N(0)$.

To illustrate the modification process we use examples. To make things easier, we only show two neighbouring pairs of leaves. Suppose that the $(p+1)$th violation occurs between the leaves of pair u and pair $t = u+1$ at x_w. Note that $I_{tlp}(x_w) \neq I_{ulp}(x_w)$, as otherwise, either (b) or (c) of lemma 2 is true. In case $I_{tlp}(x_w) > I_{ulp}(x_w)$, swap u and t. Now, we have $I_{tlp}(x_w) < I_{ulp}(x_w)$. *In the following, we refer to these u and t values as the u and t of Algorithm TONGUEANDGROOVE.* From lemma 2 and the fact that a violation has occurred, it follows that $I_{trp}(x_w) < I_{urp}(x_w)$. To remove this tongue-and-groove constraint violation, we modify (I_{tlp}, I_{trp}). The other profiles of $N(p)$ are not modified.

The new plan for pair t, $(I_{tl(p+1)}, I_{tr(p+1)})$ is as defined below. If $I_{ulp}(x_w) - I_{tlp}(x_w) \leqslant I_{urp}(x_w) - I_{trp}(x_w)$, then

$$I_{tl(p+1)}(x) = \begin{cases} I_{tlp}(x) & x_0 \leqslant x < x_w \\ I_{tlp}(x) + \Delta I & x_w \leqslant x \leqslant x_m \end{cases} \tag{8}$$

where $\Delta I = I_{ulp}(x_w) - I_{tlp}(x_w)$. $I_{tr(p+1)}(x) = I_{tl(p+1)}(x) - I_t(x)$, where $I_t(x)$ is the target profile to be delivered by the leaf pair t.

Otherwise,

$$I_{tr(p+1)}(x) = \begin{cases} I_{trp}(x) & x_0 \leqslant x < x_w \\ I_{trp}(x) + \Delta I' & x_w \leqslant x \leqslant x_m \end{cases} \tag{9}$$

where $\Delta I' = I_{urp}(x_w) - I_{trp}(x_w)$. $I_{tl(p+1)}(x) = I_{tr(p+1)}(x) + I_t(x)$, where $I_t(x)$ is the target profile to be delivered by the leaf pair t.

The former case is illustrated in figure 6 and the latter is illustrated in figure 7. Note that our strategy for plan modification is similar to that used by van Santvoort and Heijmen (1996) to eliminate a tongue-and-groove violation for dynamic multileaf collimator plans.

Since $(I_{tl(p+1)}, I_{tr(p+1)})$ differs from (I_{tlp}, I_{trp}) for $x \geqslant x_w$ there is a possibility that $N(p+1)$ is involved in tongue-and-groove violations for $x \geqslant x_w$. Since none of the other leaf profiles are changed from those of $N(p)$ no tongue-and-groove constraint violations are

3737

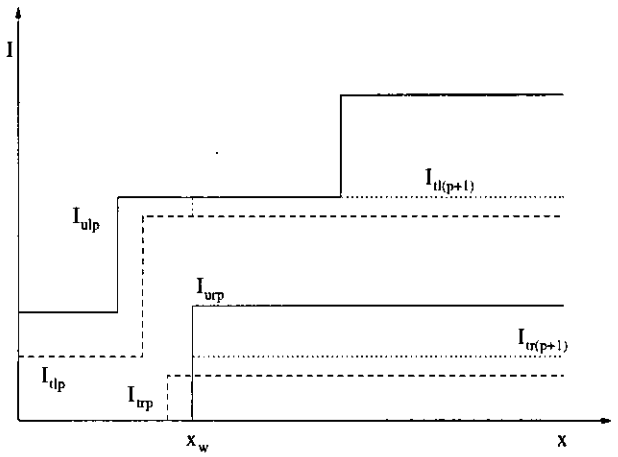


Figure 6. Tongue-and-groove constraint violation: case 1.

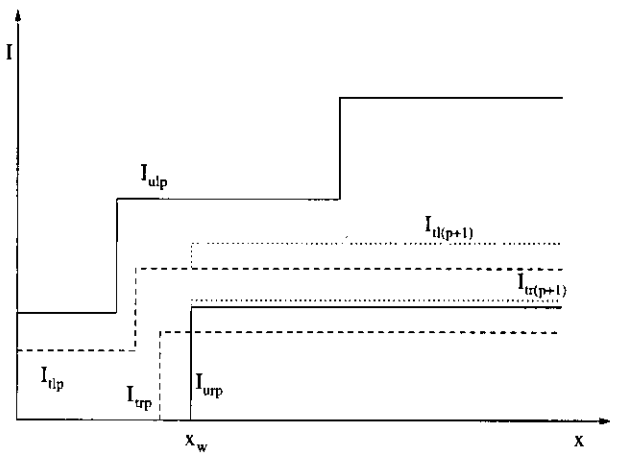


Figure 7. Tongue-and-groove constraint violation: case 2 (close parallel dotted and solid line segments overlap, they have been drawn with a small separation to enhance readability).

possible in $N(p+1)$ for $x < x_w$. One may also verify that since I_{tl0} and I_{tr0} are non-decreasing functions of x, so also are I_{tlp} and I_{trp}, $p > 0$.

Lemma 3. *Let* $F = ((I'_{1l}, I'_{1r}), (I'_{2l}, I'_{2r}), \ldots, (I'_{nl}, I'_{nr}))$ *be any unidirectional schedule for the desired profile that satisfies the tongue-and-groove constraint. Let* $S(p)$ *be the following assertions.*

(a) $I'_{il}(x) \geqslant I_{ilp}(x) \qquad 0 \leqslant i \leqslant n \quad x_0 \leqslant x \leqslant x_m$
(b) $I'_{ir}(x) \geqslant I_{irp}(x) \qquad 0 \leqslant i \leqslant n \quad x_0 \leqslant x \leqslant x_m$

$S(p)$ *is true for* $p \geqslant 0$.

 3758

Proof. The reader is referred to Kamath *et al* (2003b) for the proof. □

2.3.2. Elimination of tongue-and-groove effect and interdigitation. As we have pointed out, the elimination of tongue-and-groove constraint violations does not guarantee elimination of interdigitation constraint violations. Therefore the schedule generated by Algorithm TONGUEANDGROOVE may not be free of interdigitation violations. The algorithm we propose for obtaining schedules that simultaneously satisfy both constraints, Algorithm TONGUEANDGROOVE-ID, is similar to Algorithm TONGUEANDGROOVE. The only difference between the two algorithms lies in the definition of the constraint condition. To be precise we make the following definition.

Definition 1. *A unidirectional schedule is said to satisfy the tongue-and-groove-id constraint if*

(a) $I_{tr}(x_i) \leqslant I_{(t+1)r}(x_i) \leqslant I_{(t+1)l}(x_i) \leqslant I_{tl}(x_i)$, *or*
(b) $I_{(t+1)r}(x_i) \leqslant I_{tr}(x_i) \leqslant I_{tl}(x_i) \leqslant I_{(t+1)l}(x_i)$,

for $0 \leqslant i \leqslant m, 1 \leqslant t < n$.

The only difference between this constraint and the tongue-and-groove constraint is that this constraint enforces condition *(a)* or *(b)* above to be true at all sample points x_i including those at which $I_t(x_i) = 0$ and/or $I_{t+1}(x_i) = 0$.

Lemma 4. *A schedule satisfies the tongue-and-groove-id constraint iff it satisfies the tongue-and-groove constraint and the interdigitation constraint.*

Proof. It is obvious that the tongue-and-groove-id constraint subsumes the tongue-and-groove constraint. If a schedule has a violation of the interdigitation constraint, $\exists i, t, I_{(t+1)l}(x_i) < I_{tr}(x_i)$ or $I_{tl}(x_i) < I_{(t+1)r}(x_i)$. From definition 1, it follows that schedules that satisfy the tongue-and-groove-id constraint do not violate the interdigitation constraint. Therefore a schedule that satisfies the tongue-and-groove-id constraint satisfies the tongue-and-groove constraint and the interdigitation constraint.

For the other direction of the proof, consider a schedule O that satisfies the tongue-and-groove constraint and the interdigitation constraint. From the fact that O satisfies the tongue-and-groove constraint and from lemma 2 and definition 1, it only remains to be proved that for schedule O,

(a) $I_{tr}(x_i) \leqslant I_{(t+1)r}(x_i) \leqslant I_{(t+1)l}(x_i) \leqslant I_{tl}(x_i)$, or
(b) $I_{(t+1)r}(x_i) \leqslant I_{tr}(x_i) \leqslant I_{tl}(x_i) \leqslant I_{(t+1)l}(x_i)$,

whenever $I_t(x_i) = 0$ or $I_{t+1}(x_i) = 0, 0 \leqslant i \leqslant m, 1 \leqslant t < n$.

When $I_t(x_i) = 0$,

$$I_{tl}(x_i) = I_{tr}(x_i). \tag{10}$$

Since O satisfies the interdigitation constraint,

$$I_{tr}(x_i) \leqslant I_{(t+1)l}(x_i) \tag{11}$$

and

$$I_{(t+1)r}(x_i) \leqslant I_{tl}(x_i). \tag{12}$$

From equations (10)–(12), we get $I_{(t+1)r}(x_i) \leqslant I_{tr}(x_i) = I_{tl}(x_i) \leqslant I_{(t+1)l}(x_i)$. So (b) is true whenever $I_t(x_i) = 0$. Similarly, (a) is true whenever $I_{t+1}(x_i) = 0$. Therefore, O satisfies the tongue-and-groove-id constraint. □

3739

Algorithm TONGUEANDGROOVE-ID

(i) $x = x_0$

(ii) While (there is a tongue-and-groove-id violation) do

(iii) Find the least x_w, $x_w \geq x$, such that there exist leaf pairs u, $u+1$, that violate the tongue-and-groove-id constraint at x_w.

(iv) Modify the schedule to eliminate the violation between leaf pairs u and $u+1$.

(v) $x = x_w$

(vi) End While

Figure 8. Obtaining a schedule under both the constraints.

Algorithm TONGUEANDGROOVE-ID finds violations of the tongue-and-groove-id constraint from left to right in exactly the same manner in which Algorithm TONGUEANDGROOVE detects tongue-and-groove violations. Also, the violations are eliminated as before, i.e., as prescribed by equations (8) and (9) and illustrated in figures 6 and 7, respectively. Algorithm TONGUEANDGROOVE-ID is shown in figure 8. All notation used in the algorithm and the related discussion in the remainder of section 2.3.2 is also the same as that used in section 2.3.1 and corresponds directly to the usage in Algorithm TONGUEANDGROOVE.

Lemma 5. *Let* $F = ((I'_{1l}, I'_{1r}), (I'_{2l}, I'_{2r}), \ldots, (I'_{nl}, I'_{nr}))$ *be any unidirectional schedule for the desired profile that satisfies the tongue-and-groove-id constraint. Let* $S(p)$*, be the following assertions.*

(a) $I'_{il}(x) \geqslant I_{ilp}(x), 0 \leqslant i \leqslant n, x_0 \leqslant x \leqslant x_m$

(b) $I'_{ir}(x) \geqslant I_{irp}(x), 0 \leqslant i \leqslant n, x_0 \leqslant x \leqslant x_m$

$S(p)$ *is true for* $p \geqslant 0$.

Proof. The reader is referred to Kamath *et al* (2003b) for the proof. □

2.4. Efficient implementation of the algorithms

In the remainder of this section we will use 'algorithm' to mean Algorithm TONGUEANDGROOVE or Algorithm TONGUEANDGROOVE-ID and 'violation' to mean tongue-and-groove constraint violation or tongue-and-groove-id constraint violation (depending on which algorithm is considered) unless explicitly mentioned.

The execution of the algorithm starts with schedule M at $x = x_0$ and sweeps to the right, eliminating violations from the schedule along the way. The modifications applied to eliminate a violation at x_w, prescribed by equations (8) and (9), modify one of the violating profiles for $x \geqslant x_w$. From the unidirectional nature of the sweep of the algorithm, it is clear that the modification of the profile for $x > x_w$ can have no consequence on violations that may occur at the point x_w. Therefore it suffices to modify the profile only at x_w at the time the violation at x_w is detected. The modification can be propagated to the right as the algorithm sweeps. This can be done by using an $(n \times m)$ matrix A that keeps track of the amount by which the profiles have been raised. $A(j, k)$ denotes the cumulative amount by which the jth leaf pair profiles have been raised at sample point x_k from the schedule M generated using Algorithm MULTIPAIR. When the algorithm has eliminated all violations at each x_w, it moves to x_{w+1} to look for possible violations. It first sets the $(w+1)$th column of the modification matrix

equal to the wth column to reflect rightward propagation of the modifications. It then looks for and eliminates violations at x_{w+1} and so on.

The process of detecting the violations at x_w merits further investigation. We show that if one carefully selects the order in which violations are detected and eliminated, the number of violations at each x_w, $0 \leqslant w \leqslant m$ will be $O(n)$.

Lemma 6. *The algorithm can be implemented such that $O(n)$ violations occur at each x_w, $0 \leqslant w \leqslant m$.*

Proof. The bound is achieved using a two-pass scheme at x_w. In pass one we check adjacent leaf pairs $(1, 2), (2, 3), \ldots, (n-1, n)$, in that order, for possible violations at x_w. In pass two, we check for violations in the reverse order, i.e., $(n-1, n), (n-2, n-1), \ldots, (1, 2)$. So each set of adjacent pairs $(i, i+1)$, $1 \leqslant i < n$ is checked exactly twice for possible violations. It is easy to see that if a violation is detected in pass one, either the profile of leaf pair i or that of leaf pair $i+1$ may be modified (raised) to eliminate the violation. However, in pass two only the profile of pair i may be modified. This is because the profile of pair i is not modified between the two times it is checked for violations with pair $i+1$. The profile of pair $i+1$, on the other hand, could have been modified between these times as a result of violations with pair $i+2$. Therefore in pass two, only i can be a candidate for t (where t is as explained in the algorithm) when pairs $(i, i+1)$ are examined. From this it also follows that when pairs $(i-1, i)$ are subsequently examined in pass two, the profile of pair i will not be modified. Since there is no violation between adjacent pairs $(1, 2), (2, 3), \ldots, (i, i+1)$ at that time and none of these pairs is ever examined again, it follows that at the end of pass two there can be no violations between pairs $(i, i+1)$, $1 \leqslant i < n$. □

Lemma 7. *For the execution of the algorithm, the time complexity is $O(nm)$.*

Proof. Follows from lemma 6 and the fact that there are m sample points. □

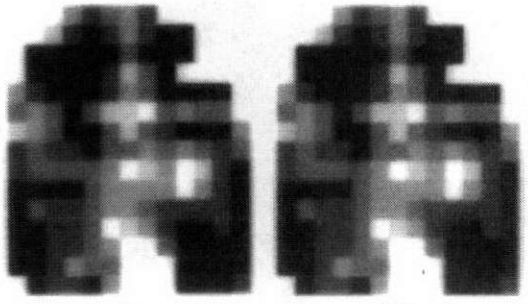
Theorem 1.

(a) Algorithms TONGUEANDGROOOVE and TONGUEANDGROOVE-ID terminate.
(b) The schedule generated by Algorithm TONGUEANDGROOVE is free of tongue-and-groove constraint violations and is optimal in therapy time for unidirectional schedules.
(c) The schedule generated by Algorithm TONGUEANDGROOVE-ID is free of interdigitation and tongue-and-groove constraint violations and is optimal in therapy time for unidirectional schedules.

Proof.

(a) Lemma 7 provides a polynomial upper bound ($O(n*m)$) on the complexity of Algorithms TONGUEANDGROOVE and TONGUEANDGROOVE-ID. The result follows from this.
(b) When Algorithm TONGUEANDGROOVE terminates, no tongue-and-groove violations remain. From this and lemma 3, it follows that the schedule generated by Algorithm TONGUEANDGROOVE is optimal in therapy time for unidirectional schedules free of tongue-and-groove violations.
(c) When Algorithm TONGUEANDGROOVE-ID terminates, no tongue-and-groove-id violations remain and from lemma 4 the final schedule satisfies the tongue-and-groove and interdigitation constraints. From this and lemma 5, it follows that the schedule generated

3741



(a) (b)

Figure 9. Film measurement of the AP field (field ID 1 in table 1) of a 7-field head and neck plan. The optimized leaf sequences were generated without (Algorithm MULTIPAIR, (a)) and with tongue-and-groove-id correction (Algorithm TONGUEANDGROOVE-ID, (b)).

Table 1. Comparison of the number of segments and MU efficiency of the three leaf sequencing algorithms (MULTIPAIR, TONGUEANDGROOVE and TONGUEANDGROOVE-ID) for seven intensity maps of a head and neck treatment plan generated from a commercial treatment planning system. The per cent increases in the number of segments and MUs for Algorithms TONGUEANDGROOVE and TONGUEANDGROOVE-ID with respect to Algorithm MULTIPAIR are also shown. The average per cent increases in the number of segments are 21% and 25%, respectively. The average per cent increases in the number of MUs are 19% and 24%, respectively.

	Field ID						
	1	2	3	4	5	6	7
MULTIPAIR							
Number of segments	11	8	13	15	14	10	10
MU efficiency	0.47	0.63	0.40	0.35	0.37	0.40	0.51
TONGUEANDGROOVE							
Number of segments	14	10	15	21	14	13	11
MU efficiency	0.37	0.51	0.35	0.25	0.37	0.40	0.40
% Segment (number) increase	27	25	15	40	0	30	10
% MU increase	26	24	15	37	0	0	28
TONGUEANDGROOVE-ID							
Number of segments	14	11	16	21	14	14	11
MU efficiency	0.37	0.47	0.33	0.25	0.37	0.37	0.37
% Segment (number) increase	27	38	23	40	0	40	10
% MU increase	26	36	22	37	0	7	38

by Algorithm TONGUEANDGROOVE-ID is optimal in therapy time for unidirectional schedules free of both types of violations. □

Theorem 2. *The schedule generated by the algorithm of van Santvoort and Heijmen (1996) is free of interdigitation and tongue-and-groove constraint violations and is optimal in therapy time for unidirectional DMLC schedules with this property.*

Proof. Similar to that of theorem 1(c). □

3. Experimental validation

The algorithms were validated on a Varian 2100 C/D with 120-leaf MLC (Varian Medical Systems, Palo Alto, CA). The intensity maps of a 7-field head and neck plan from a commercial inverse treatment planning system (CORVUS 5.0, NOMOS Corporation, Cranberry, PA) were sequenced using Algorithm MULTIPAIR, which optimizes the MU efficiency, and Algorithm TONGUEANDGROOVE-ID, which eliminates the tongue-and-groove effect and interdigitation. The intensity maps have a bixel size of 1 cm $\times$ 1 cm and a 20% intensity step. Figure 9 shows the film measurement of the fluence maps of the AP field. The tongue-and-groove effect is readily seen in figure 9(a), while it is completely eliminated in figure 9(b) using Algorithm TONGUEANDGROOVE-ID. Table 1 compares the number of segments and the MU efficiencies of all three algorithms. The MU efficiency is defined as the ratio of the maximum fluence of intensity modulated field per MU to the fluence of an open field per MU. Compared to the leaf sequences with no constraints, the consideration of tongue-and-groove correction increased both the number of segments and MUs, with an average increase of 21% and 19%, respectively, for the seven intensity maps considered here. With the additional elimination of interdigitation, the increases were 25% and 24%, respectively. Examination of all the sub-fields of the leaf sequences generated with Algorithm TONGUEANDGROOVE-ID verified that no interdigitation constraint has been violated.

4. Conclusions

We have described a mathematical formalism and rigorous proofs of leaf sequencing algorithms for segmental multileaf collimation, which maximize MU efficiency while completely eliminating the tongue-and-groove underdosage. Even though it has been shown that for a multiple field IMRT plan ($\geqslant$5), the tongue-and-groove effect on the IMRT dose distribution is clinically insignificant (Deng *et al* 2001) due to the smearing effect of individual fields, yet it still can be problematic for a small number of fields and for the patient setup with minimal uncertainty. Compared to the unconstrained leaf sequencing algorithms, the presented methods yield leaf sequences which decrease the MU efficiency a little. But they completely overcome tongue-and-groove underdosages. One of the methods also eliminates leaf interdigitation. Most importantly, mathematical proofs show that these algorithms are optimal in MU efficiency for unidirectional schedules.

Acknowledgments

This work was supported, in part, by the National Library of Medicine under grant LM06659-03. We are grateful to Mr Arvind Kumar for making the observation that a two-pass implementation of the algorithm resolved all tongue-and-groove violations for test sets. This observation motivated us to formulate lemma 6.

References

Chui C-S, LoSasso T and Spirou S 1994 Dose calculation for photon beams with intensity modulation generated by dynamic jaw or multileaf collimations *Med. Phys.* **21** 1237–44

Convery D J and Webb S 1998 Generation of discrete beam-intensity modulation by dynamic multileaf collimation under minimum leaf separation constraints *Phys. Med. Biol.* **43** 2521–38

Deng J, Pawlicki T, Chen Y, Li J, Jiang S and Ma C-M 2001 The MLC tongue-and-groove effect on IMRT dose distributions *Phys. Med. Biol.* **46** 1039–60

Dirkx M L P, Heijmen B J M and van Santvoort J P C 1998 Leaf trajectory calculation for dynamic multileaf collimation to realize optimized fluence profiles *Phys. Med. Biol.* **43** 1171–84

Galvin J, Smith A and Lally B 1993a Characterization of a multileaf collimator system *Int. J. Radiat. Oncol. Biol. Phys.* **25** 181–92

Galvin J M, Chen X-G and Smith R M 1993b Combining multileaf fields to modulate fluence distributions *Int. J. Radiat. Oncol. Biol. Phys.* **27** 697–705

Kamath S, Sahni S, Li J, Palta J and Ranka S 2003a Leaf sequencing algorithms for segmented multileaf collimation *Phys. Med. Biol.* **48** 307–24

Kamath S, Sahni S, Palta J, Ranka S and Li J 2003b *Optimal Leaf Sequencing with Elimination of Tongue-and-Groove Underdosage* http://www.cise.ufl.edu/~sahni/papers/tng.pdf

Mohan R 1995 Field shaping for three-dimensional conformal radiation therapy and multileaf collimation *Seminars Radiat. Oncol.* **5** 86–99

Sykes R J and Williams P C 1998 An experimental investigation of the tongue and groove effect for the Philips multileaf collimator *Phys. Med. Biol.* **43** 3157–65

van Santvoort J P C and Heijmen B J M 1996 Dynamic multileaf collimation without 'tongue-and-groove' underdosage effects *Phys. Med. Biol.* **41** 2091–105

Wang X, Spirou S, LoSasso T, Stein J, Chui C and Mohan R 1996 Dosimetric verification of intensity modulated fields *Med. Phys.* **23** 317–28

Webb S, Bortfeld T, Stein J and Convery D 1997 The effect of stair-step leaf transmission on the 'tongue-and-groove problem' in dynamic radiotherapy with a multileaf collimator *Phys. Med. Biol.* **42** 595–602

Xia P and Verhey L J 1998 Multileaf collimator leaf sequencing algorithm for intensity modulated beams with multiple static segments *Med. Phys.* **25** 1424–34

AAPM REPORT NO. 28

QUALITY ASSURANCE METHODS AND PHANTOMS FOR MAGNETIC RESONANCE IMAGING

Published for the
American Association of Physicists in Medicine
by the American Institute of Physics

AAPM REPORT NO. 28

QUALITY ASSURANCE METHODS AND PHANTOMS FOR MAGNETIC RESONANCE IMAGING[†]

REPORT OF
TASK GROUP NO. 1
NUCLEAR MAGNETIC RESONANCE COMMITTEE*

AAPM

Members

Ronald R. Price (Task Group Chairman)
Leon Axel
Tommie Morgan
Robert Newman
William Perman
Nicholas Schneiders
Mark Selikson
Michael L. Wood[‡]
Stephen R. Thomas

[†]Reprinted from MEDICAL PHYSICS, Volume 17, Issue 2, 1990

*Ronald R. Price, Nuclear Magnetic Resonance Committee Chairman
Stephen R. Thomas, Past Committee Chairman

[‡]Michael L. Wood, Current Task Group Chairman

May 1990

Published for the
American Association of Physicists in Medicine
by the American Institute of Physics

Further copies of this report may be obtained from:

American Institute of Physics
c/o AIDC
64 Depot Road
Colchester, Vermont 05446

(1-800-445-6638)

International Standard Book Number: O-8831 8-800-7
International Standard Serial Number: 0271-7344

Published by the American Institute of Physics, Inc.
335 East 45 Street, New York, NY 10017

Printed in the United States of America

3747

Quality assurance methods and phantoms for magnetic resonance imaging: Report of AAPM nuclear magnetic resonance Task Group No. 1[a]

Ronald R. Price, Leon Axel, Tommie Morgan, Robert Newman, William Perman, Nicholas Schneiders, Mark Selikson, Michael Wood, and Stephen R. Thomas
AAPM Task Group No. 1

(Received 6 September 1989; accepted for publication 30 October 1989)

I. INTRODUCTION

The purpose of this document is to describe a standard set of test procedures which can be used to evaluate the performance of clinical magnetic resonance imaging systems. These procedures and tests are not intended to establish absolute performance standards but are rather intended to provide methods which can be used as part of a routine quality assurance program. It is the position of this document that the purpose of a quality assurance program is to detect changes in system performance relative to an established baseline.

This document also includes recommendations for acceptable magnetic resonance imaging (MRI) phantom materials, phantom designs, and analysis procedures. Specific image parameters described in ths document are: resonance frequency, signal-to-noise, image uniformity, spatial linearity, spatial resolution, slice thickness, slice position/separation, and phase related image artifacts. It is recognized that this set is not exhaustive and does not include procedures for assessing all possible image parameters, and similarly it is also recognized that there are acceptable methods other than those presented for measuring many of these parameters. The proposed set, however, is considered to be adequate for monitoring the sensitivity and geometric characteristics of clinical nuclear magnetic resonance (NMR) imaging systems.

The proposed set does not include specific procedures for monitoring the accuracy or precision of T1, T2 or proton density. Since at the present time, there are no commonly accepted standard methods for determining T1, T2 and proton density from image data and the assessment of these parameters is not currently a part of clinical practice, we have chosen not to include these test procedures until more knowledge on their utility and measurement is available.

The National Electrical Manufacturers Association (NEMA) is acknowledged for their assistance in the development of the sections on field uniformity and signal-to-noise. The specifications for these two parameters are consistent with the NEMA specifications wherever possible under the requirement that the procedures are applicable for use in a practical quality assurance program. The American College of Radiology (ACR) Subcommittee on Magnetic Resonance (MR) Nomenclature and Phantom Development under the MR Committee on Imaging Technology and Equipment is acknowledged for its persistent and careful review over the several years that this document has been under development.

II. PHANTOM MATERIALS

The primary considerations which dictate the choice of phantom materials for use in quality assurance phantoms are: chemical and thermal stability, the absence of significant chemical shifts, appropriate T1, T2 and proton density values which are within the biological range. As will be noted later, coil loading is an important consideration when assessing signal-to-noise. Other considerations generally relate to convenience and practicality: convenience by matching the T1 of the material to an acceptable TR which does not require an exceedingly long scan time and practicality by not choosing a T2 value shorter that some instruments can accommodate. Care should be taken to avoid the use of colored plastics or other container materials which possess significantly different magnetic susceptibility from the filler material.

At each operating field strength, it is recommended that the chosen NMR material should exhibit the following characteristics:

$$100\text{ms} < T1 < 1200\text{ms}$$
$$50\text{ms} < T2 < 400\text{ms}$$
$$\text{proton density} \gg H_2O \text{ density}$$

Numerous materials have been used successfully as NMR phantom agents. These have primarily consisted of oils and water solutions of various paramagnetic ions. For reference, listed in Table I are approximate relaxation times for mixtures of 1,2 propanediol in distilled water (Ref. 1) and three paramagnetic agents [$CuSO_4$(Ref. 2)) $NiCl_2$(Ref. 2) and $MnCl_2$(Ref. 3)] at 20 MHz (0.5 T). It should be noted that relaxation times are temperature and field-strength dependent.

The relaxation rates (inverse of relaxation times) are approximately linear with ion concentration.

For all measurements, scan conditions should be carefully recorded. Scan conditions should include: pulse sequence and scan timing parameters (TE, TI, TR), flip angle, field-of-view and matrix size, coil, phantom and phantom material, slice number and thickness, center-to-center spacing, number of acquisitions, rf power settings and any image processing which may have been used. All phantoms should be centered at the magnet isocenter unless otherwide specified.

Action criteria listed in this document arc for reference only. Absolute values of quality control parameters are ma-

 0094-2405/90/020287-09S01.20

TABLE I. Approximate relaxation times of NMR phantom materials.

Agent	Concentration	T1	T2
$CuSO_4$	1-25 mM	860-40 ms	625-38 ms
$NiCl_2$	1-25 mM	806-59 ms	763-66 ms
Propanediol	0-100%	2134-217 ms	485-72 ms
$MnCl_2$	0.1-1 mM	982-132 ms	...

chine dependent and as a result, make it impossible to specify action criteria which can be applied universally to all systems. Specific action criteria must be arrived at individually for each system installation in cooperation with the user and instrument manufacturer.

III. RESONANCE FREQUENCY

A. Definition

The resonance frequency is defined as that rf frequency f which matches the static B-field (B_o) according to the Larmor equation:

$$f = \frac{\gamma}{2\pi} B_o,$$

g is the gyromagnetic ratio for the nuclei under study. For protons, the Larmor frequency is 42.58 MHz/T, e.g., for a 1.5-T system, the resonance frequency should be 63.87 MHz.

B. Factors affecting resonance frequency

Prior to the performance of any imaging protocol, it is essential that the operator verify that the system is on resonance. Most vendors insist upon a resonance frequency check each time the imaging system is turned on. Resonance frequency checks are most important for mobile units and some resistive magnet systems which undergo frequent ramping of the magnetic field. Changes in the resonance frequency reflect changes in the static B-field. Changes in the static B-field may be due to superconductor "run down" (typically on the order of 1 ppm/day, e.g., ~60 Hz/day at 1.5 T), changes in current density due to thermal or mechanical effects, shim-coil changes or effects due to external ferromagnetic materials.

The effects of off-resonance operation relate primarily to system sensitivity and are manifest as a reduction in image signal-to-noise. Secondary effects are reflected in image linearity due to the summation of the image gradients with the inconsistent static B-field value.

It is recommended that a resonance frequency check be performed prior to quality assurance measurement and each time a different phantom is used.

C. Methods of Measurement

1. Phantom

The phantom which is used most often for resonance frequency checks in a uniform signal producing cylinder and is the same phantom that is used for the signal-to-noise measurements. The phantom is positioned in the center of the magnet (with all gradient fields turned off) and the rf frequency is adjusted by controlling the rf synthesizer center frequency to achieve maximum signal. Some resistive systems may also allow adjustment of the magnet current to alter the magnetic field strength to achieve resonance. Most vendors will provide a specific user protocol for resonance frequency adjustment and some may be completely automated. Resonance frequency should be recorded daily for trend analysis.

2. Scan conditions

No scan is required for this measurement.

3. Analysis

Resonance frequency value is recorded for comparison to previous determinations.

D. Action criterion

Values of resonance frequency should generally not deviate by more than 50 ppm between successive daily measurements. Action should also be taken any time there is a significant change in trend.

IV. SIGNAL-TO-NOISE RATIO

A. Definition

The signal is defined as the mean pixel value within the region-of-interest minus any pixel offset. Noise is defined as the random variations in pixel intensity. Images with obvious artifacts are not suitable for signal-to-noise determinations.

B. Factors affecting signal-to-noise ratio

Factors contributing to variations in signal-to-noise ratio include: (i) general system calibration (resonance frequency, flip angles, etc.) (ii) gain, (iii) coil tuning, (iv) rf shielding, (v) coil loading, (vi) image processing, and (vii) scan parameters.

C. Methods of measurement

1. Phantom

The phantom should consist of a uniform signal producing material which has a minimum dimension in the image plane of at least 10 cm or 80% of the field-of-view, whichever is larger (Fig. 1). For single slice measurements, the phantom should have a dimension in the direction of the slice selection which is at least twice the maximum slice thickness being used. For multislice acquisitions, the phantom length should be at least as long as the volume being imaged, plus two maximum slice thicknesses. The phantom may be either circular or rectangular in cross section. When using large volume fluid-filled phantoms, it should be recognized that thermal and mechanically induced motions can introduce artifacts.

The standard phantom specified here is to be filled with nonconducting material, and thus is not intended to simulate the clinical situation. The unloaded coil allows the evaluation of system noise which is the parameter of interest. In a

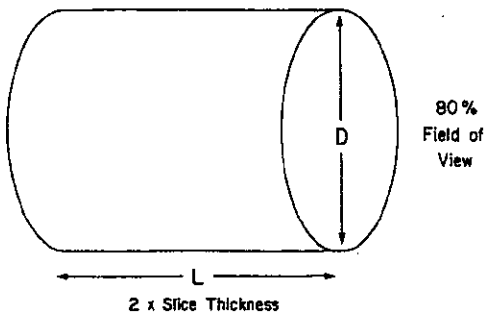


FIG. 1. The phantom used for resonance frequency, signal-to-noise ratio and image uniformity is typically composed of a uniform signal producing material. The minimum dimension *(D)* in the image plane should be at least 10 cm or 80% of the image field-of-view, whichever is larger. The length *(L)* in the slice selection direction should be at least twice the maximum slice thickness for single-slice measurements. For multislice measurements, *L* should be at least as long as the volume being imaged (slice separation-x number of slices), plus a thickness equal to twice the maximum slice thickness.

clinical scan, it is recognized that the patient is the dominant source of noise. In order to approximate the clinical situation, the coil must be electrically loaded by using an appropriate filler material or by some other means, whereby the electrical properties of the body are simulated.

Worthy of note is that the NEMA standard for signal-to-noise does specify loading for the measurement and thus differs from the signal-to-noise measurement specified in this document. It should also be noted that systems with certain high-Q coils may not be tunable under unloaded conditions.

2. Scan conditions

Any typical (usually multislice) acquisition may be used.

3. Analysis

The signal is measured using a region of interest (ROI) which contains at least 100 pixels or 10% of the area of the signal producing material, whichever is greater. The ROI should be positioned in the center of the image and should not include any obvious artifacts. The signal is the mean value of the pixel intensity in the ROI minus any offset. (An indication of the existence of an image intensity offset may be gained from an examination of intensity values from ROI's taken over nonsignal producing portions of a phantom. Specific offset values should be obtained from the system manufacturer). The noise is the standard deviation derived from the same ROI. The signal-to-noise ratio (SNR) is then calculated.

An alternative method of SNR measurement is to acquire two consecutive scans with identical scan parameters which are subsequently subtracted. This method specifically excludes the effects of low-frequency image variations. A third pixel-by-pixel difference image (image 3) is then created. The signal is defined as above using either of the original unsubtracted images. The noise is defined as the standard deviation (SD) derived from using the same ROI on the subtracted image (image 3).

The calculated signal-to-noise is as follows:

$$\mathrm{SNR} = \frac{S\sqrt{2}}{\mathrm{SD}}$$

The factor of $\sqrt{2}$ is required because the SD is derived from the subtraction image rather than from one of the original images.[4]

D. Action criterion

An action criterion can not be given since SNR results are only applicable to the specific system, phantom and scan conditions being used. It is important to re-emphasize that the signal and noise measurements are dependent on essentially all scan parameters and test conditions. SNR should be normalized to voxel size for comparison.

V. IMAGE UNIFORMITY

A. Definition

Image uniformity refers to the ability of the MR imaging system to produce a constant signal response throughout the scanned volume when the object being imaged has homogeneous MR characteristics.

B. Factors affecting image uniformity

Parameters contributing to the image nonuniformity include: (i) static-field inhomogeneities, (ii) rf field non-uniformity, (iii) eddy currents, (iv) gradient pulse calibration, and (v) image processing.

C. Methods of measurement

1. Phantom

The characteristics of the phantom used for image uniformity evaluation are identical to the characteristics of the phantom used for signal-to-noise determination (Sec. IV). To prevent rf penetration effects, the filler material should be non-conducting.

Nonuniformities resulting from rf penetration effects may be evaluated by scanning a phantom which has been filled with a conductive solution such as normal saline. Due to partitioning in the body, penetration effects observed in a scan of a saline-filled phantom will not necessarily predict penetration effects which would be found in human scans.

2. Scan conditions

Any typical multislice acquisition may be used provided the signal-to-noise ratio is sufficiently large so that it does not affect the uniformity measurement. Adequate signal-to-noise ratio may be insured by either increasing the number of acquisitions or by applying a low-pass smoothing filter. In practice, it has been found that a signal-to-noise ratio of 80:1 or greater will yield good results.

3. Analysis

For pixels within a centered geometric area which encloses approximately 75% of the phantom area, the maxi-

mum (S_{max}) and minimum (S_{min}) values are determined. Care should be taken to not include edge artifacts in the ROI. A span Δ and midrange value $\overline{S}$ are calculated as follows:

$$\Delta = \frac{S_{max} - S_{min}}{2}$$

$$\overline{S} = \frac{S_{max} + S_{min}}{2}.$$

The relationship for calculating integral uniformity (U_Σ) is

$$U_\Sigma = \left[1 - \frac{\Delta}{S}\right] \times 100\% = \left[1 - \frac{(S_{max} - S_{min})}{(S_{max} + S_{min})}\right] \times 100\%.$$

Perfect integral uniformity using this relationship is when $U_\Sigma = 100\%$.

In some cases (e.g., low-field imaging) signal-to-noise may be a limiting factor in the measurement of image uniformity. To help minimize the effect of noise on the measurement the image may be convolved with a nine-point low-pass filter $h(m_1,m_2)$. The filtered image is given by

$$s(n_1,n_2) = \frac{1}{W} \sum_{m_1=-1}^{1} \sum_{m_2=-1}^{1} h(m_1,m_2) s(n_1 - m_1, n_2 - m_2),$$

where $n1$, $n2$ cover the range of the image.

The filter kernel is

$$\frac{1}{W}\begin{bmatrix} 1 & 2 & 1 \\ 2 & 4 & 2 \\ 1 & 2 & 1 \end{bmatrix} = \begin{bmatrix} h(-1,-1) & h(-1,0) & h(-1,1) \\ h(0,-1) & h(0,0) & h(0,1) \\ h(1,-1) & h(1,0) & h(1,1) \end{bmatrix},$$

and represents the product of two raised cosines in the frequency domain. The weighting factor W is given by

$$W = \sum_{m_1=-1}^{1} \sum_{m_2=-1}^{1} h(m_1,m_2) = 16,$$

and is used to normalize the dc response of the filter in the frequency domain to unity. This filter has a 3-dB cutoff spatial frequency contour which very closely approximates a circle of radius 0.364 π in normalized coordinates. It is the two-dimensional equivalent of the Hanning filter. The above filter gives a gain in the signal-to-noise ratio of 2.4.

D. Action criterion

For a 20-cm field-of-view or less, the integral uniformity should be typically 80% or better. It should be realized that for larger fields-of-view, the uniformity may deteriorate. Image uniformity in the above context is not defined for surface coils.

VI. SPATIAL LINEARITY

A. Definition

Spatial linearity is a term used to describe the degree of geometrical distortion present in images produced by any imaging system. Geometrical distortion can refer to either displacement of displayed points within an image relative to their known location, or improper scaling of the distance between points anywhere within the image.

B. Factors affecting spatial linearity

The primary factors which introduce geometrical distortion in NMR imaging are: (i) inhomogeneity of the main magnetic field and; (ii) nonlinear magnetic field gradients.

C. Methods of measurement

1. Phantom

Variability is best observed over the largest field-of-view. The phantom to be used to measure spatial linearity should occupy at least 60% of the largest field-of-view and consist of a regular array of objects (holes, grooves, rods, or tubes) of known dimensions and spacing, and the phantom filled with signal producing material. The objects within the array should be of a size in which the location can be measured and spaced in a regular pattern (typically every 1-2 cm). The dimensional positioning error of the objects within the array, due to finite pixel size, should be < 10% of the linearity specification. Figure 2 provides an illustration of two possible patterns which could be used to evaluate spatial linearity.

2. Scan conditions

Consideration should be given to determining the spatial linearity for a typical multislice acquisition with the largest available image matrix to maximize spatial resolution.

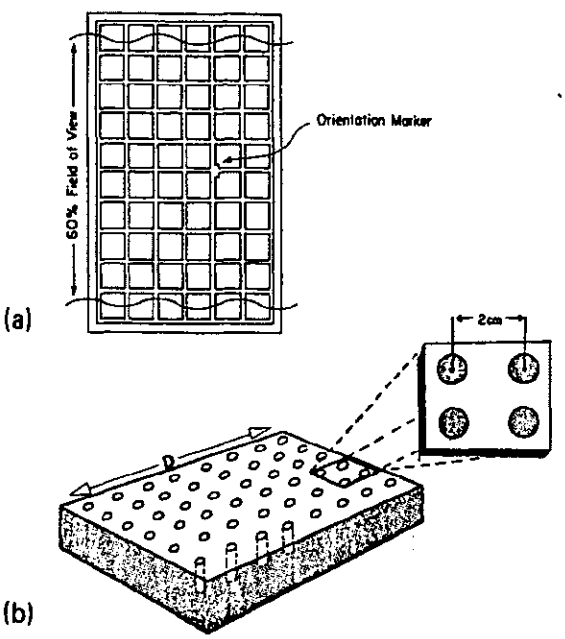


FIG. 2. The phantom used for spatial linearity should have a minimum dimension (D) in the image plane of at least 60% of the largest possible image field-of-view. The thickness of the phantom should be at least twice the maximum slice thickness for single-slice measurements and two slice thicknesses, plus the image volume length for multislice measurements. Two possible phantom designs are (a) orthogonal grooves in an acrylic plate of (b) an orthogonal array of holes drilled in an acrylic plate. Orientation markers are recommended.

3751

Since NMR imaging is inherently a volumetric imaging technique, the evaluation should be performed for each orthogonal plane to define the useful imaging volume. This can be done either by using a specially designed phantom for multislice image acquisition or by using a single-slice phantom placed at different locations and in the three orthogonal orientations. Spatial linearity is not expected to depend significantly on image timing parameters such as TE, TR and the number of signal acquisitions.

If oblique planes are frequently used, consideration should be given to the inclusion of linearity measurements for oblique planes, as well as the orthogonal planes.

3. *Analysis*

Percent distortion is defined as

$$\frac{\text{true dimension} - \text{observed dimension}}{\text{true dimension}} \times 100\%.$$

Distortion measurement may be performed between any two points within the field-of-view, provided that pixel-resolution is not a significant source of error. It is recommended that the true dimension be greater than 10 pixels. Preliminary considerations by the NEMA task group to specify image distortion have centered on the use of a cylindrical phantom in which several measured diameters are compared to the known diameter. Specification in terms of the maximum deviation (maximum-minimum) expressed as a percent of the known diameter is also under consideration.

Spatial linearity measurements performed directly on the image processing unit will provide information about the MR imaging system alone. Measurements can also be performed upon filmed images and will provide combined performance information about the MR imager, as well as the video and filming systems.

D. Action criterion

Percent distortions in the spatial linearity (when measured over a 25 cm or greater field-of-view) are generally considered acceptable if they are < 5%.

VII. HIGH-CONTRAST SPATIAL RESOLUTION

A. Definition

High-contrast spatial resolution is a measure of the capacity of an imaging system to show separation of objects when there is no significant noise contribution. High-contrast spatial resolution for MRI systems is typically limited by acquisition matrix pixel size (field-of-view divided by the sampling in x or y). The acquisition matrix pixel size should not be confused with the display matrix pixel size in which pixel interpolation or replication may have occurred.

Traditionally, resoltuion has been quantified by the point spread function (PSF), line spread function (LSF), or modulation transfer function (MTF); however, these methods are not practical for routine quality assurance measurements on MRI systems. Therefore, a visual evaluation of test objects will be used.

B. Factors affecting resolution

Factors contributing to high-contrast resolution include: field-of-view (determined by gradient strength and sampling period), acquisition matrix and reconstruction filters.

C. Methods of measurement

1. *Phantom*

Useful spatial resolution phantoms for visual evaluation may be composed of either bar patterns or hole (or rod) arrays. Array signal-producing elements may be either round or rectangular in cross section. The patterns consist of alternating signal producing and nonsignal producing areas set apart from each other by a width equal to the bar's or hole's width, i.e., center-to-center spacing is twice the diameter. Square bar patterns offer an advantage over round cross-section (hole) patterns in that the smallest resolvable array element can be related to resolution in terms of line-pairs per millimeter.

A typical phantom (Fig. 3) may consist of five signal producing elements and four spaces with element sizes of 5, 3, 2, 1.5, 1.25, 1.00, 0.75, and 0.50 mm, although additional increments may be used. The dimension in the slice selection direction (length) should be at least twice the slice thickness, i.e., 20 mm length for 10 mm slice thickness.

2. *Scan conditions*

Any typical multislice acquisition may be used provided it incorporates an appropriate slice thickness (nominal 5-10 mm) to insure an adequate signal-to-noise. The phantom should be aligned perpendicular to the scan plane and locat-

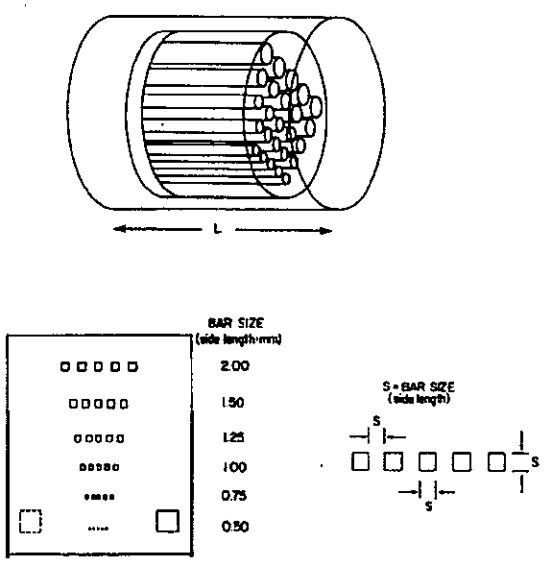


FIG. 3. High-contrast resolution phantoms may be composed of either bar patterns or hole arrays. Bars or holes should have center-to-center spacings *(S)* equal to twice the hole diameter or bar dimension. The length *(L)* of the phantom should be at least twice the maximum slice thickness. Bars derived from grooves in an acrylic sheet may be preferred due to construction difficulties.

ed at the isocenter and should be rotated at 45° within the image plane to combine the resolution from both the phase and frequency encoding directions. In order to determine the resolution in the phase and frequency encoding directions independently, two scans will be required in which the phantom resolution elements are aligned along each axis separately and then scanned.

3. *Analysis*

The image will be evaluated visually. Image analysis consists of viewing the image to determine the smallest resolvable array element (magnification may be used if desired). For an array to be resolved, all five elements and four spaces must be displayed as separate and distinct when viewed with the narrowest window width. The window level should be adjusted for optimum visualization. Resolution is expressed as the size of the smallest resolvable array element or its equivalent in lp/mm when square bar patterns are used.

D. Action criterion

The high-constrast resolution should remain constant for repeated measurements under the same scan conditions and should be equal to the pixel size. For example for a 25.6 cm field-of-view with a 256x256 acquisitions matrix, the resolution should be 1 mm.

VIII. SLICE THICKNESS

A. Definition

Slice thickness is defined as the full width at half-maximum (FWHM) of a slice profile. The full width at tenth-maximum (FWTM) is an additional descriptor of the slice profile. The slice profile is defined as the response of the magnetic resonance imaging system to a point source as it moves through the plane of the reconstruction at that point.

B. Factors affecting slice thickness

(i) Gradient field nonuniformity, (ii) rf field nonuniformity, (iii) nonuniform static magnetic field, (iv) noncoplanar slice selection pulses between excitation and readout, (v) TR/T1 ratio, and (vi) rf pulse shape and stimulated echoes.

C. Methods of measurement

1. *Phantoms*

Several phantoms can be used to evaluate slice thickness, most of which utilize some variant of an inclined surface (plane, cone or spiral). A typical phantom is the crossed high signal ramps.

High signal ramp (HSR) phantoms generally consist of opposing ramp pairs oriented at a fixed angle (Θ) [Fig. 4(a)] with respect to one another. The HSR's should be thin (ideally infinitesimally thin) in order to quantify the slice profile accurately. Because of the low signal in the image imposed by the small volume of signal-producing material in a thin ramp, averages of pixel values across the width of the ramp may be needed to generate a slice profile with an acceptable SNR. As thinner (< 3 mm) slice thicknesses are evaluated, it is necessary to increase ramp angle and to decrease ramp thickness. In general, the thickness of a (90°) HSR oriented at 45° respect to the image plane should be < 20% of the slice profile FWHM (i.e., 5-mm slice needs a 1-mm ramp) to get a measurement with < 20% error.

An alternative method which is particularly useful for evaluating thin slices is the use of the slice selection echo method.[5] A standard selective 90° and 180° pulse sequence may be used together with a readout gradient oriented along the slice selection direction. The Fourier transform of the resulting echo gives a picture of the slice profile. The strength of the readout gradient is needed to translate the frequency axis to actual spatial dimensions.

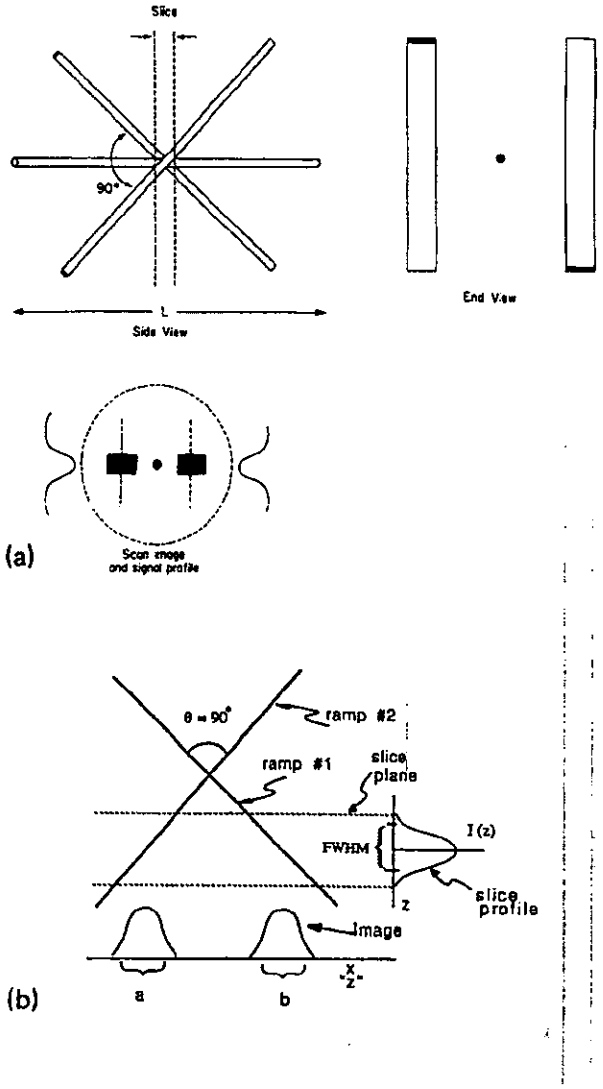


FIG. 4. (a) A typical slice-thickness phantom consists of two crossed thin ramps. A ramp crossing angle of 90° yields an angle of 45° between the ramp and the image plane. Ramp thickness should typically be < 20% of the slice thickness being evaluated. Phantom length *(L)* should be greater than twice the maximum slice thickness. An alignment rod passing between the two ramps defines the point where the two ramps cross. When the slice is properly aligned through the intersection of the ramps the images of the ramps and rod image will all be aligned. (b) The slice sensitivity profile will be directly proportional to the image intensity profiles if the image plane is perpendicular to the alignment rod. By using the geometric mean of the two profiles ($\sqrt{ab}$) correct FWHM values are obtained even with image plane misalignment.

2. Scan conditions

Any typical multislice acquisition may be used provided TR is greater than 3T1 of the filler material and the highest pixel resolution is used. Slice thickness should be measured both centrally and peripherally within an image and at both central (magnet isocenter) and offset slice locations.

3. Analysis

Slice thickness (FWHM, FWTM) : In the resultant image, the signal level is read out across the ramp on a pixel-by-pixel basis along a line-of-interest oriented orthogonally to the ramp width dimension. As noted previously, to assure adequate S/N, it may be necessary to either use multiple excitations or several line profiles. The FWHM or FWTM parameters should be determined for each of the dual ramps.

The general equation for the FWHM from imaging opposed high signal ramps (relative angle Θ) oriented at any angle with respect to the image plane is

$$\text{FWHM} = \frac{(a+b)\cos\Theta + \sqrt{(a+b)^2\cos^2\Theta + 4ab\sin^2\Theta}}{2\sin\Theta}$$

where a and b refer to the measured FWHM (FWTM) of the intensity profiles for ramp 1 and ramp 2, respectively [Fig. 4(b)].[6]

For the case of $\Theta = 90°$, the equation simplifies to:

$$\text{FWHM} = \sqrt{ab}$$

D. Action criterion

Assuring adequate measurement accuracy, the measured value of slice thickness should generally agree with the indicated slice thickness within ± 1 mm for slice thicknesses > 5 mm.

IX. SLICE POSITION/SEPARATION

A. Definition

Slice position (offset) is the absolute location of the midpoint of the FWHM of the slice profile. Slice separation is the distance between any two slice positions. Slice locations are indicated by external positioning devices or by the selected interslice spacing.

B. Factors affecting slice position/separation

(i) Misalignment of positioning devices, (ii) gradient field nonuniformity, (iii) B1 nonuniformity, (iv) noncoplanar slice selection pulses, and (v) static magnetic field.

C. Methods of measurement

1. Phantoms

In general, the same phantom used for slice thickness measurements [Fig. 4(a)] may also be used for slice position/separation determinations, with the provision that the phantom contains reference pins and external scribed marks for orientation, centering, and reference to the external positioning devices. An inclined surface, with a known pitch, when imaged at different locations will produce images which will be displaced relative to a reference in direct proportion to the slice location and the pitch of the surface.

2. Scan conditions

Any typical acquisition is suitable for slice position/separation determinations.

3. Analysis

The midpoint of the FWHM of the slice profile in the image of interest is determined. (Fig. 5). The distance *(D)* from the profile midpoint to a landmark (alignment rod) which remains stationary from slice-to-slice (parallel to the slice selection direction) is measured and related to the slice position *(O)*. For a 45° ramp, the distance from a centered reference pin to the slice profile midpoint will be equal to the slice distance from the magnet isocenter if the phantom is accurately positioned with the crossover point of the ramps located at the isocenter. For any relative ramp angle (Θ) the slice offset position *(O)* will be given by

$$O = D / \tan(\Theta / 2).$$

All measurements should be made along the line defined by the magnet isocenter and the centers of the imaging planes.

D. Action criterion

Comparison of external position marker should generally agree with the actual slice position within ± 2 mm. Slice separation disagreement should typically be < 20% of the total slice separation or ± 1 mm, whichever is greater.

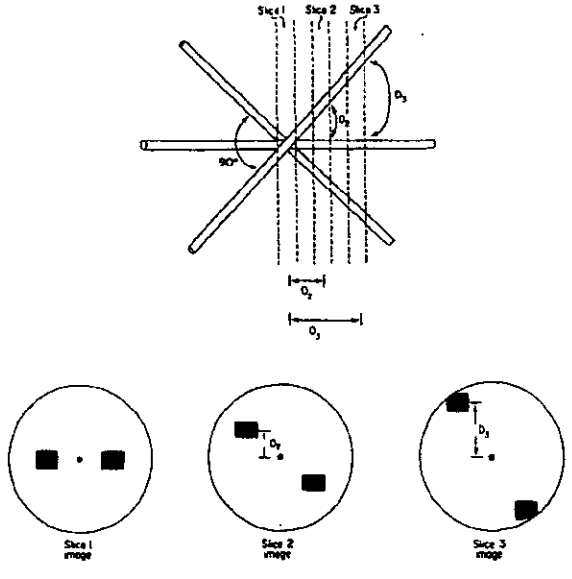


FIG. 5. Slice position (offset) and separation may be determined using a phantom similar to the slice-thickness phantom (Fig. 4). Slices taken at different locations (0,) (slices 1-3) in a multislice sequence will produce images of the ramps which are progressively further from the alignment rod *(D)*. The distance D is measured in the image and then related to the true slice location *(O)* from the isocenter.

X. IMAGE ARTIFACTS

A. Definition

Phase related errors are defined in terms of inappropriate (either increased or decreased) image signal at specified spatial locations. Generally, these artifacts are characterized by increased signal intensity in areas which are known to contain no signal producing material. Commonly called "ghosts", errors in the application of phase-encoding gradients for imaging and errors in both rf transmit and receive quadrature phase, result in unique ghost artifacts. A "dc-offset" error is defined here as high-intensity or low-intensity pixels at the center of the image matrix due to improper scaling of low-frequency components (typically dc) in the Fourier transformation of the NMR time-domain signal.

B. Factors affecting phase related artifacts

(i) Phase encoding gradient instability, (ii) quadrature phase maladjustment in the synthesis of slice selective rf pulses (transmit error), and (iii) improper quadrature phase decoding on receive.

C. Methods of measurement

1. Phantom

A typical phantom design is illustrated in Fig. 6. It consists of a single signal producing cylinder (2-5 cm) located at an asymmetric location, typically on the periphery of the field-of-view at a 45° orientation. The phantom thickness should be approximately twice the slice thickness being used. Orientation markers are particularly beneficial for this phantom.

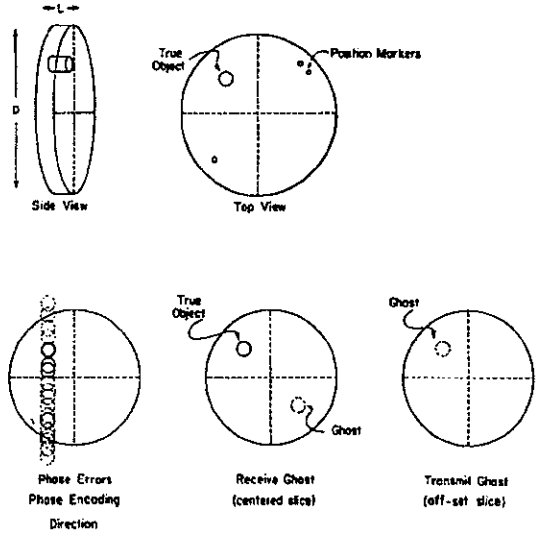


FIG. 6. A typical phantom for quadrature error detection consists of a single signal producing cylinder (labeled "true-object") located at an asymmetric location, e.g., at the periphery of the field-of-view at a 45° orientation. The size of the cylinder is not critical and may be as large as 2-5 cm in diameter. Marker sources are important for orientation information. Phantom diameter *(D)* should be at least 10 cm. Phantom thickness *(L)* should be at least two times the maximum slice thickness.

2. Scan conditions

Any typical multislice sequence may be used. Separate scans must be made to assess both transmit and receive errors if a phantom similar to the phantom in Fig. 5 is used. More complex volume phantoms may be designed in which both transmit and receive errors may be assessed with a single-scan sequence. The scan for assessing receive quadrature errors is made with the phantom placed at the magnet isocenter with the central slice of the multislice sequence passing through the phantom. The same scan may be used to assess both dc-offset and phase encoding errors. The scan for assessing transmit quadrature errors is made with the phantom placed at a convenient offset slice position (typically ± 5 cm from the isocenter slice) with the center slice passing through the magnet isocenter and an offset slice passing through the phantom.

3. Analysis

a. Phase-encoding errors. Phase-encoding ghosts will appear as multiple images (possibly smeared into a column) originating at the true object position but displaced along the phase-encoding axis of the image (perpendicular to the frequency encoding direction). The presence of these characteristic ghost images will generally identify the two axes; however, the orientations should be verified by the manufacturer or operator's manual. Regions-of-interest values are taken from both the true image and the brightest ghost image. The magnitude of the error *(E)* is quantified by expressing the ghost ROI value *(G)* as a percent of the true ROI *(T)*:

$$E = \frac{(T-G)}{T} \times 100\%.$$

b. dc-offset errors. dc-offset errors typically appear as a single bright pixel (sometimes as a dark pixel if overflow or processing has occurred) at the center of the image matrix. The existence of this error is assessed visually.

c. Receive quadrature errors. Receive quadrature ghosts will be evaluated using the central slice of the multislice sequence acquire with the phantom at the isocenter. Receive ghosts will appear upside down and reversed from the true signal producing object (object in the upper left-hand corner will appear as a ghost in the lower right-hand corner). Regions-of-interest values are taken from both the true image and the ghost image. The receive quadrature error *(E)* is quantified by expressing the ghost ROI value *(G)* as a percent of the true ROI *(T)*.

$$E = \frac{(T-G)}{T} \times 100\%.$$

d. Transmit quadrature errors. Transmit quadrature ghosts are evaluated using images acquired in multislice mode in which the phantom is placed at a location offset from the isocenter. A transmit ghost appears in the slice located in the opposite offset direction at a distance equal to the distance at which the true object is located from the isocenter (mirror image from the isocenter). The ghost and true object image will be located at the same relative positions in their respective images. For example, a true object

3755

located in the upper left-hand corner at a distance of + 5 cm from the isocenter will produce a transmit quadrature ghost in the upper left-hand corner of the image at - 5 cm. ROI's taken over the true object and the ghost are used to determine the percent error *(E)*.

$$E = \frac{(T-G)}{T} \times 100\%.$$

D. Action criterion

Phase related errors should typically be < 5% of the true signal value. dc-offset errors should not be present in images from a properly functioning system.

Further information on MRI quality assurance methods and phantoms may be found in the scientific literature (7-38).

[a]Task Group No. 1 is part of the AAPM Nuclear Magnetic Resonance Committee, Stephen R. Thomas, Chairman, during development of document (current Chairman Ronald R. Price). This document has been co-sponsored by the American College of Radiology, MR Committee on Imaging Technology and Equipment, Alexander R. Margulis, Chairman.

[1]R. R. Price, J. A. Patton, J. J. Erickson, et al., "Concepts of Quality Assurance and Phantom Design for NMR Systems," Medical Physics Monograph *No.* 14, *NMR in Medicine: The Instrumentation and Clinical Application*, edited by S. R. Thomas and R. L. Dixon, (American Institute of Physics, NY, 1985), p. 414.

[2]M. Bucciolini, L. Cimolo and R. Renzi, Med. Phys. 13,298-303 (1986).

[3]P. T. Beale, S. R. Amtey, and S. R. Kasturi *NMR Data Handbook for Biomedical Applications* (Pergamon Press, New York, 1984).

[4]William R. Hendee, *Medical Radiation Physics*. "Accumulation and Analysis of Nuclear Data", (Year Book Medical, Chicago, 1979), Chap. 12.

[5]D. I. Hoult, "NMR Imaging Techniques," 40, 132-138 (1984).

[6]D. R. White, R. D. Speller, and P. M. Taylor, "Evaluating Performance Characteristics in Computed Tomography," Br. J. Radiol. 54, 221-231 (1981).

[7]J. M. S. Hutchinson, R. J. Sutherland, and J. R. Mallard "NMR Imaging: Image Recovery Under Magnetic Fields with Large Non-Uniformities," J. Phys. E. Sci. Instrum. 11,217-221 (1978).

[8]Ching-Ming Lai, "Reconstructing NMR Images from Projections under Inhomogeneous Magnetic Field and Non-linear Field Gradients, Phys. Med. Biol. 28, 925-938 (1983).

[9]V. M. Runge, C. T. Johnson, and F. W. Smith, "Phantoms for Magnetic Resonance Imaging," Noninvas. Med. Imag. 1, 49-60 (1984).

[10]I. R. Young, D. J. Bryant, I. A. Payne, "Variations in slice shape and absorption as artifacts in the determination of tissue parameters in NMR Imaging", Magnetic Resonance in Medicine, 2, 355-389 (1985).

[11]M. O'Donnell and W. A. Edelstein, "NMR Imaging in the Presence of Magnetic Field Inhomogeneities and Gradient Field Non-linearities," Med. Phys. 12, 20-26 (1985).

[12]R. A. Lerski, K. Straughan, J. S. Orr, "Calibration of Proton Density Measurements in Nuclear Magnetic Resonance Imaging," Phys. Med. Biol. 271-276 (1986).

[13]W. A. Edelstein, G. H. Glover, C. J. Hardy, R. W. Redington, "The Intrinsic Signal-to-Noise Ratio in NMR Imaging," Magnetic Resonance in Medicine 3, 604-618 (1986).

[14]L. Brateman, L. W. Jennings, R. L. Nunnally, et al., "Evaluations of Magnetic Resonance Imaging Parameters with Simple Phantoms," Med. Phys. 13, 441-448 (1986).

[15]D. W. McRobbie, R. A. Lerski, K. Straughan, "Investigation of Slice Characteristics in Nuclear Magnetic Resonance Imaging. Phys. Med. Biol. 31, 613-626 (1986).

[16]E. M. Bellon, E. M. Haacke, P. E. Coleman, "M. R. Artifacts: A Review," Am. J. of Roent. 147, 1271-1281 (1986).

[17]E. Pusey D. D. Stark, R. B. Lufkin, "Magnetic Resonance Imaging Artifacts: Mechanism and Clinical Significant," Radiographics 6, 891-911 (1986).

[18]R. K. Breger, F. E. W. Wehrli, H. C. Charles, "Reproducibility of Relaxation and Spin Density Parameters in Phantoms and the Human Brain Measured by MR Imaging at 1.5T," Mag. Res. in Med. 3, 649-662 (1986).

[19]I. Mano, H. Goshima, M. Namba, "New Polyvinyl Alcohol Gel Material for MRI Phantoms," Mag. Res. in Med. 3, 921-926 (1986).

[20]J. R. Kowles and J. A. Markisz, "Upholding MR Image Quality can be a Complex but Profitable Pursuit," Diagnostic Imaging 125-130 (1987).

[21]R. M. Henkelman, and M. J. Bronskill, "Artifacts in Magnetic Resonance Imaging," *Reviews of Magnetic Resonance in Medicine: Special Issue*, (Pergamon, New York, 1987), Vol. 2.

[22]J. A. Patton, M. V. Kulkarni, J. K. Craig, "Techniques, Pitfalls and Artifacts in Magnetic Resonance Imaging," Radiographics 7, 505-519 (1987).

[23]J. E. Gray, "Section Thickness and Contiguity Phantom for MR Imaging," Radiology 164, 193-197 (1987).

[24]"Identification and Characterization of Biological Tissue by NMR. Concerted Research Project of the European Economic Community," edited by John C. Gore and Francis W. Smith. Special Editorial Mag. Res. Imag. 6, 171-222 (1988).

[25]"IV. Protocols and Test Objects for the Assessment of MRI Equipment: EEC Concerted Research Project," edited by John C. Gore and Francis W. Smith, Mag. Res. Imag. 6, 195-199 (1988).

[26]R. A. Lerski, D. W. McRobbie, K. Straughan, P. M. Walker, J. D. de Certaines and A. M. Bernard, "V. Multi-Center Trial with Protocols and Prototype Test Objects for the Assessment of MRI Equipment," Mag. Res. Imag. 6, 201-214 (1988).

[27]P. Walker, R. A. Lerski, DeVre-Mathur, J. Binet and F. Yane, "VI. Preparation of Agarose Gels as Reference Substances for NMR Relaxation Time Measurements," Mag. Res. Imag., 6, 215-222 (1988).

[28]J. C. Blechinger, B. C. Madsen and G. R. Frank, "Tissue Mimicking Gelatin-Agar Gels for Use in Magnetic Resonance Imaging Phantoms," Med. Phys., 15,629-636 (1988).

[29]W. A. Edelstein, P. A. Bottomley and L. M. Pfeiffer, A Signal-to-Noise Calibration Procedure for NMR Imaging Systems," Med. Phys. 11, 180-185 (1984).

[30]C. W. Coffey, R. Taylor, C. T. Umstead, "A Slice Geometry Phantom for Cross-Sectional Tomographic Imagers," Med. Phys. 16, 273-278 (1989).

[31]M. Chui, D. Blakesley, S. Mohapata, "Test Method for MR Image Slice Profile," J. Comp. Assist. Tomogr. 9, 1150-1152 (1985).

[32]M. Grey and C. W. Coffey: Method for Evaluating Image Quality in Magnetic Resonance Imaging, Radiol. Technol. 58, 339 (1987).

[33]M. Selikson and T. Fearon, "Averaging Error in NMR Slice Profile Measurements," Magn. Reson. Med. 7, 280 (1988).

[34]B. R. Condon, J. Patterson, D. Wyper, et al., "Image Nonuniformity in Magnetic Resonance Imaging: Its Magnitude and Methods for Correction," Br. J. Radiol., 60, 83-87 (1987).

[35]D. W. McRobbie, "Quality Assurance and Specification Measurements in NMR Imaging," in *Quality Assurance in Medical Imaging*. (The Institute of Physics, Bristol, 1986), pp. 49-66.

[36]M. M. Corell, D. O. Mearshen, P. L. Carson, et al., "Automated Analysis of Multiple Performance Characteristics in Magnetic Resonance Imaging Systems," Med. Phys. 13, 815-823 (1986).

[37]M. E. Masterson et al., "Accuracy and Reproducibility of Image Derived Relaxation Times," Med. Phys. 16, 229-233 (1989).

[38]"MRI: Acceptance Testing and Quality Control," *Proceedings of AAPM Symposium*. Winston-Salem, North Carolina, edited by Robert L. Dixon, (Medical Physics Publishing Corporation, Madison, Wisconsin 1988).

An Image Space Algorithm for Morphological Contour Interpolation

William Barrett, Eric Mortensen, and David Taylor
Department of Computer Science
Brigham Young University
Provo, Utah

e-mail: barrett@cs.byu.edu
Telephone: (801)-378-7430

Abstract

An image space algorithm for morphological interpolation between contours is presented. Image space interpolation avoids the need to represent or store contour data using intermediate data structures. The algorithm makes use of basic morphological transforms such as dilation and erosion and interimage operations such as XOR and union. Morphological interpolation is applied successfully to a variety of synthetic contours as well as naturally occurring contours such as those found in medical images or topographic maps [17]. The algorithm interpolates between nested, overlapping, nonoverlapping, or branching contours in a general way although nonoverlapping or minimally overlapping contours require initial registration. The algorithm is particularly appropriate for generation of digital elevation maps or whenever the original contour data is derived from a regular sampling grid. Image space morphological interpolation exploits pipeline architectures allowing simultaneous generation of interpolated contour values while making essential use of neighboring contour morphology. In addition, there is a logarithmic gain in the number of interpolated points when processing a contour interval exhaustively.

keywords: contour interpolation, morphological transforms, parallel, height grid, DTM, DEM, cartography

1. Introduction

Many real world objects are effectively and succinctly represented by contours. For example, geologic terrain surfaces can be represented by nested, usually nonintersecting, isocontours found in topographic maps. Isocontours often are used to generate Digital Terrain Models (DTMs) or (discrete) Digital Elevation Models (DEMs) in automated cartography [1-4]. Contours also may be extracted from and used to represent closed three-dimensional objects such as medical anatomy. For example, contours of anatomical boundaries may be detected automatically from serial transaxial cross sections in CT or MRI scans [5]. In the case of closed three-dimensional objects, contours from two spatially adjacent slices frequently intersect or overlap when superimposed.

Because it may not be economically or physically practical to densely sample the object of interest, contours usually provide only a sparse representation of the object(s) from which they were extracted. As a result, interpolation schemes [6-10, 18] often are necessary to recover the original three-dimensional surface geometry. Thus, connection of, or equivalently, interpolation between contours is a general problem in computer graphics. The most significant advancements in approaching this problem have come through algorithms which address the correct mapping or correlation of contours points at one level with those at an adjoining level [6, 18].

Some of the greatest difficulties associated with a general solution to the problem of contour interpolation arise due to topological changes and/or overlap between adjacent contours, such as when contours differ in both position and number from one level to an adjacent level. Such is the case when there is a natural branching of the surface geometry. Even without branching, if there is a striking disparity in contour shape or position between two adjacent levels, interpolation algorithms may fail to provide a smooth transition of object geometry at intermediate levels. Thus, efficient and robust contour interpolation algorithms which make essential use of contour morphology at both the local and the global level are still needed.

A new image space contour interpolation algorithm which exploits both local and global contour morphology has been developed. The algorithm makes use of morphological transforms (such as dilation and erosion) and other image-level logical operations (AND, OR, or XOR), all of which operate directly in image space. The main idea of the algorithm is to find the midline between two contours and use it to split the intercontour space into two halves, each of which can be processed in the same way. This process is repeated recursively until the intercontour space is exhausted (i.e. filled with midlines). If the initial two contours overlap, the first midline simply passes through the point of

3757

intersection and the algorithm proceeds as usual.

When compared to existing techniques, image space morphological contour interpolation offers advantages in robustness, accuracy, and computation. These include

1. Robustness: Handles any number of contours of any shape including branching or overlapping geometries. Nonoverlapping contours must be registered.
2. Accuracy: Makes essential use of contour morphology; local shape is extracted using dilation and erosion operations while global shape is represented by midlines.
3. Computation: All operations are performed directly in image space which avoids intermediate contour representation and storage while exploiting pipeline architectures. For nested contours this results in massively parallel speedup since *all* contour intervals as well as entire families of interpolated contour

In particular, for *each* contour interval and across all the parallelism *increases* $O(2^{m-1})$ with each recursion while the number of operations decreases $O(\log_2 n)$. This is because contour intervals are essentially split in two with each iteration.

Morphological contour interpolation is very well suited to generation of height grid DEMs from discrete isocontours as is illustrated in this paper. However, the algorithm is also extensible and applicable to other 3D objects whose contours originate *on the grid* such as heart contours from CT scans [5].

A brief introduction to mathematical morphology is given in Section 2. This is followed in Section 3 with a definition of and a distinction between the midline and the medial axis of a region. Section 4 presents the algorithm for morphological contour interpolation followed by results from both simulated and real world contours in Section 5. Section 6 contains a summary of algorithm features with suggestions for future work.

2. Mathematical Morphology

One of the strengths of mathematical morphology lies in its ability to decompose complex shapes into their meaningful parts. In fact, some morphological transforms result directly in structures, such as the medial axis, which have powerful intrinsic shape-describing content. The intent here is to present briefly some morphological operations which are integral to the problem of contour interpolation. For a more detailed treatment of mathematical morphology see references [11-12].

2.1 Dilation and Erosion

The most basic morphological transforms are dilation and erosion. Dilation and erosion transforms exist for both binary and grayscale images. We first define binary dilation, r_b, and erosion, s_b. Let X be a binary-1 object such as shown in the image in Figure 1a. (Black dots indicate binary-1 pixels; empty squares have value 0.) Let B be a structuring element of binary-1 pixels such as shown in Figure 1b. (A structuring element is similar to a convolution kernel in image processing.) Let B_x be the translation of B so that the origin of B is located at position x in the object image. The binary dilation of X by B ($X\ r_b\ B$, Figure 1c) is obtained by passing B over the object image and ORing B to the (initially 0) output image whenever the origin of B is over a binary-1 pixel $x \in X$. Formally,

$$X\ r_b\ B = \{x \mid B_x \cap X \neq \emptyset\} \qquad (1)$$

The binary erosion of an object X by B, ($X\ s_b\ B$) is obtained by passing B over the object image and plotting the origin of B in the (initially 0) output image whenever B is completely contained in X. Formally,

$$X\ s_b\ B = \{x \mid B_x \subset X\} \qquad (2)$$

If X were represented by the binary-1 pixels in Figure 1c, binary erosion of X by B would result in the object in Figure 1a. Thus, binary erosion is the dual of binary dilation.

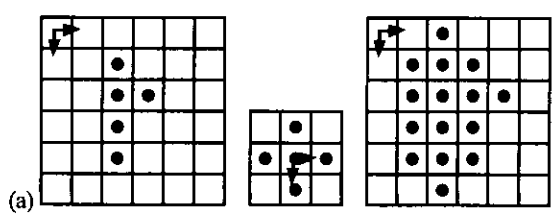


Fig. 1. (a) Binary-1 object pixels X = ● (b) Structuring element, B = ● (c) $X\ r_b\ B$ = ●. ↱ denotes origin.

Grayscale erosion and dilation can be considered an extension of the binary case. However, rather than being defined only in terms of binary set operations, grayscale erosion and dilation are defined in terms of minimum and maximum over grayscale image pixels I covered by a grayscale structuring element, G_x, with its origin translated to position x in I. Specifically, grayscale dilation of I by G ($I\ r_g\ G$), is defined as

3758

$$I \, r_g \, G = \{x \mid x = \max_{z \in G_x \cap I} (I(z) + G_x(z)\} \quad (3)$$

where I(z) are the pixel values in image I corresponding to values of $G_x(z)$ with its origin at x. Algorithmically, for each translation of G in image I, the pixel value x corresponding to the origin of G_x is calculated as follows for grayscale dilation:

Grayscale Dilation

Input: Grayscale image I, grayscale structuring element, G

Output: Dilated grayscale image I

for each pixel x in I {only process pixels
  for each pixel z in $G_x \cap I$ covered by G}
    I (z) ← $I(z) + G_x(z)$ {add I value to corresponding G value}
  I (x) ← max [I'(z) in $G_x \cap I$] {output max in I }

In contrast, grayscale erosion of I by G ($I \, s_g \, G$), uses minimization over the intersect region. Specifically, ($I \, s_g \, G$) is defined as

$$I \, s_g \, G = \{x \mid x = \min_{z \in G_x \cap I} (I(z) - G_x(z)\} \quad (4)$$

and can be computed with the appropriate changes in the dilation algorithm above. In the trivial case, the values of the structuring element G are all zero and add no bias to either the minimization or the maximization. Grayscale dilation of a simple image (Figure 2a) by a 4-connected grayscale structuring element composed only of zero values (Figure 2b) results in the image in Figure 2c, (ignoring values outside of the indicated region). Conversely, grayscale erosion of the image in Figure 2c would result in the original image in Figure 2a. This simple illustration is provided because the structuring element in Figure 2b is the one used for morphological contour interpolation described in the algorithm below.

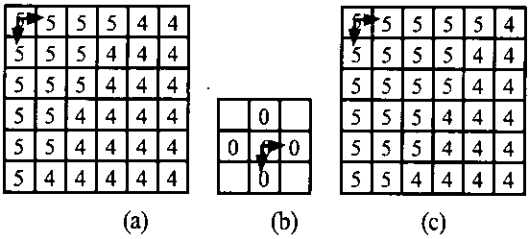


Fig. 2. (a) Grayscale image I (b) Structuring element, G (c) X r_g G. ↱ denotes origin.

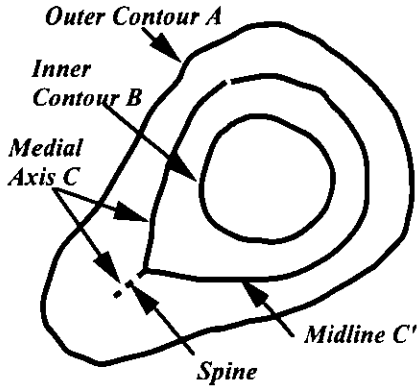


Figure 3. Medial axis, ***C***, of region bounded by outside contour ***A*** and inside contour ***B***. The medial axis includes the spine(s) (dashed) while the midline, ***C'***, does not.

2.2 Medial-Axis versus Medial Line

The medial axis [15] of a two-dimensional space-filling region, R, is defined as the locus of points, $x \in R$ which have two or more points on the boundary of R which are equidistant from x (Figure 3). If the distance, d from x to the boundary is saved with each medial axis point, the region can be regenerated from the union of all disks of radius d centered at their respective medial axis points, x. Thus, the medial axis provides intrinsic and powerful global shape description capability. Note that the medial axis as used for shape regeneration is not necessarily connected. In this application the medial axis itself, rather than shape recovery, is our goal and it will be convenient for us to depict the medial axis as a connected line.

Figure 3 shows the medial axis, ***C***, of the region defined by the intercontour space between two contours ***A*** and ***B***. Note that the medial axis contains an extension referred to here as a spine. We define the medial line or midline, ***C'***, to be the medial axis without the spine. For interpolation purposes, the medial line is preferable to the medial-axis since use of the medial axis (with spine(s)) results in an undesirable webbing artifact when used in the interpolation algorithm described below.

3. Morphological Contour Interpolation

The objective of the morphological contour interpolation algorithm which follows is to identify the height or elevation of points between contours initially labeled by height, z, thereby producing a continuous interpolated grid of contour values which defines explicitly the discrete surface represented by the original contour lines. The basic idea of the algorithm is to

expand (dilate) contours into the intercontour space until they collide. The collision front defines the medial line of the intercontour space. The medial line is labeled with the average of the original two contour labels and the process is repeated until all intercontour space pixels are labeled. This process is illustrated in Figure 4. The algorithm is given below.

Morphological Interpolation Algorithm.

Input: Labeled contours C_i (intensity = elevation)

Output: Interpolated height grid A

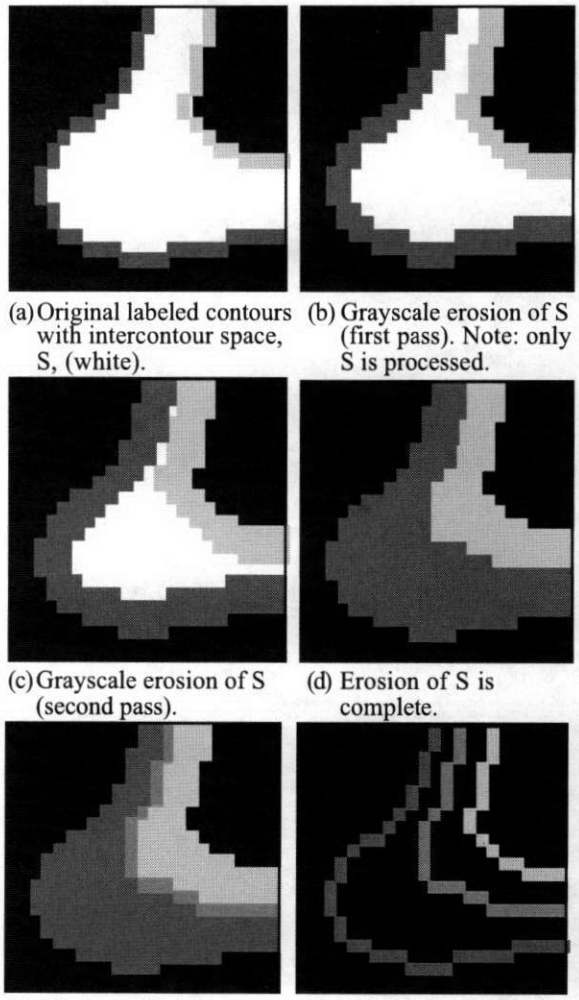
Structures:

- G 4-connected gray-scale dilation/erosion structuring element (Fig. 2b)
- A Accumulator Image (init with C_i, else zero)
- T, U, V, W, X - Work Images
- M maximum label/height

```
while  min(A) = 0 do
   W ← A                      {T, W = A with inter-
   T, W[0] ← M                contour spaces set to M}
   repeat
      W ← W[M] sg G           {Grayscale erode W until
   until max(W) < M           all M pixels are gone}

   U ← (W rg G) ∩ T           {Dilate & mask to overlap}
   V ← W XOR U                {Overlap V=medial line(s)}
   for all V > 0
      X ← M                   {X ← medial line mask}
   W ← [(U+W) ∩ X]/2          {Get medial line height}
   A ← A ∪ W                  {A ← new medial pixels}
```

Input to the algorithm consists of the original contours C_i all labeled by height. The algorithm terminates when all regions in the accumulator array A are filled with a connected set of interpolated values. The algorithm begins by copying contour lines (initialized in A) into a work buffer W and mapping intercontour pixels to high values for subsequent erosion. It then erodes the contour lines into the intercontour space until the eroded lines meet. This is almost equivalent to dilating the contours into the intercontour space with the exception that when a choice between a larger or smaller value must be made, the smaller value is chosen. This is done because the next step favors the higher pixel values and therefore chooses a medial line closer to the center than does using dilation in both instances. After the intercontour space is filled, W is dilated once and placed in a temporary buffer U. The overlap (i.e.difference) between the two buffers contains the medial lines which



(a) Original labeled contours with intercontour space, S, (white).
(b) Grayscale erosion of S (first pass). Note: only S is processed.
(c) Grayscale erosion of S (second pass).
(d) Erosion of S is complete.
(e) Dilation of intercontour space XORed with (d) gives overlap = medial line.
(f) Label for medial line = average of original contours.
(g) Intercontour spaces are relabeled for next iteration.

Figure 4. (a) Original contours labeled by height. Intercontour space S is labeled white. (b) Grayscale erosion of S erodes S and dilates original contours in the direction of S since only S is processed. (c) Original contours begin to meet with second erosion of S. (d) Erosion of S is complete. (e) Position of medial line is at collision front in figure (d). (f) Medial line is output to accumulator array = interpolated grid. Label = average of original contour labels. (g) Relabeling of new intercontour spaces allows process to be repeated until entire intercontour space is filled with interpolated values.

are extracted using an XOR operation and stored in work image V. The elevations are then computed and the medial lines are added to the accumulator.

It is worth pointing out that all interpolated values are calculated simultaneously for several families of pixels (i.e. contours) since the morphological operations are pipelined through all pixels in the image, targeting specifically the intercontour space pixels.

The process is illustrated in Figure 5 for two simple nested contours. Figure 5 also shows three medial lines resulting from the first two iterations. The interpolated height grid is shown in Figure 6. Figure 7 shows the height grid rendered using an algorithm for polygonalization of the discrete surface [17] (GI '94). The polygon rendering shows some roughness at the base and the top due to the discrete nature of the height grid and the corresponding polygonal approximation. This could be overcome by finer (perhaps) subpixel polygonalization of the discrete surface. However, that is not the objective of this paper; the polygon rendering is included here and in subsequent examples to demonstrate the surface geometry associated with the original contours.

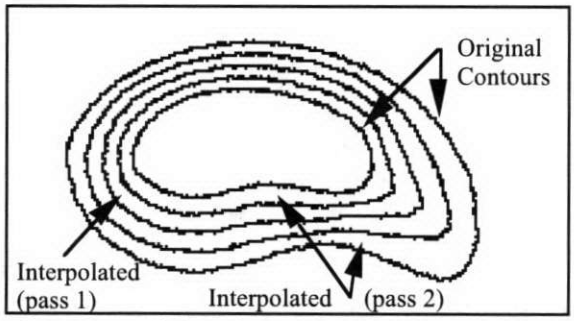


Fig. 5. Interpolation of 3 medial lines (2 passes) from two simple nested contours.

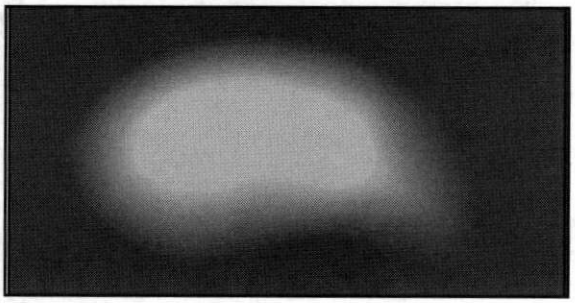


Fig. 6. Complete densely interpolated height grid.

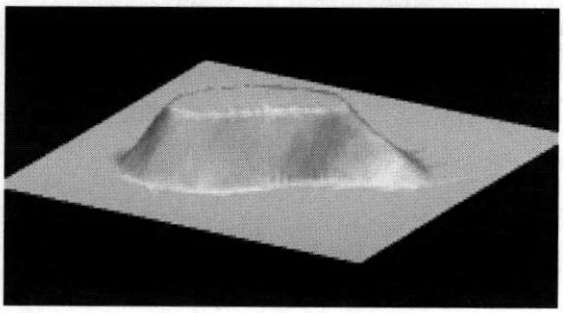


Fig. 7. Polygonal rendering of height grid in Fig. 6.

Figures 8 and 9 illustrate application of the algorithm to multiple nested contours with the corresponding surface rendering. Note the "ghosts" associated with the original contours and their respective levels. This illustrates the independent processing of each contour interval since each yield a different slope based on the (x,y) distance as well as the difference in height (z) between contours. If the medial line is identical to the medial axis, the slope is linear. If not, (i.e. spines exist) the slope or surface between two contours has negative curvature in the region of the spine, as illustrated in the top third of Figure 10. (Perhaps this is appropriate for terrain.)

One of the strong features of the algorithm is that branching contours are handled automatically without any modification to the algorithm. This is illustrated for a simple set of nested contours in Figures 10 and 11. The height of the saddle between the two contours is a function of the distance between contours and the distance to the surrounding contour. As with slopes in general this need not be linear. In fact, the depth of the saddle or the pitch of the slope could be attenuated simply through a lookup table.

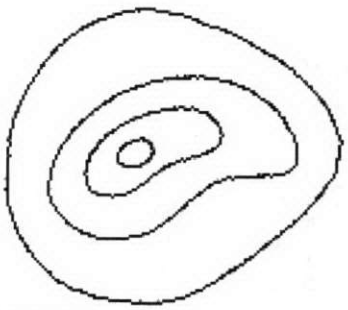


Fig. 8. Multiple nested contours.

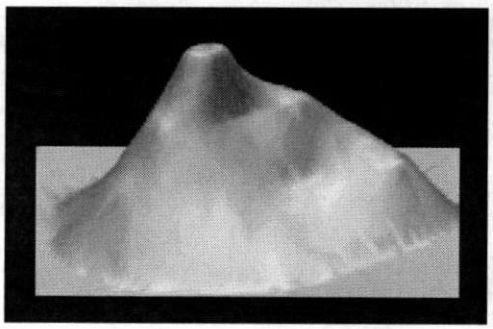


Fig. 9. Polygonal rendering for Fig. 8.

The algorithm also handles overlapping contours in a general way. Figure 12 shows two overlapping contours (black) with the interpolated medial line resulting from the first iteration (gray). In this case the initial contours are simply taken "as is." That is, no effort was made to register or align the contours. Figure 13 shows two overlapping contours which were registered by their center of gravity prior to interpolation. Registration is appropriate for minimally overlapping contours and essential if the contours do not overlap at all. The resulting interpolated contour (gray) is superimposed by offsetting it half way between the corresponding centers of gravity. The algorithm also produces reasonable results for contours which overlap and branch as evidenced in Figure 14. The indentations in the interpolated (gray) contour correspond to and agree with the branching contours (B) and become more pronounced as subsequent interpolated contours between the gray line and B gravitate towards B.

4. Results and Discussion

Figures 15-17 demonstrate application of the algorithm to a set of contours extracted from a topographic map. The height grid (DEM) in Figure 16 was created by interpolating between the isocontour lines in Figure 15. The corresponding surface rendering in Figure 17 demonstrates that the algorithm performs well on real world data as well. Figure 18 is a height grid of Mount Rundle (near Banff, of course) computed from a topographic map of the region. Polygonalization and rendering for Figure 18 can be seen in [17].

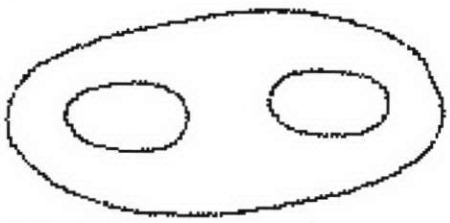


Fig. 10. Nested branching contours.

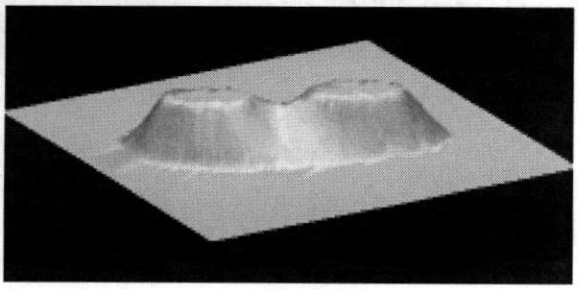


Fig. 11. Polygonal rendering for branching contours.

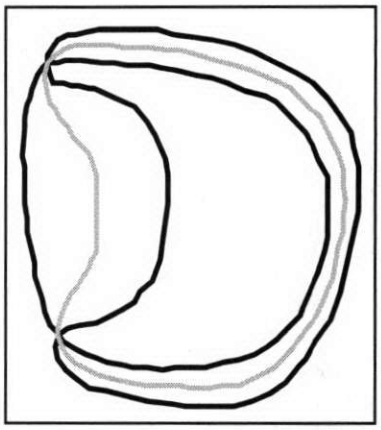


Fig. 12. Overlapping contours (black). Interpolated medial line (gray).

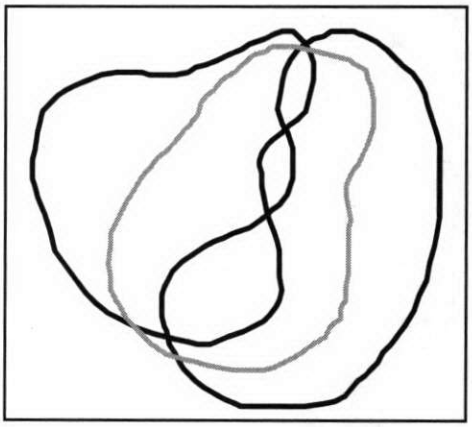


Fig. 13. Overlapping contours registered to produce interpolated medial line (gray).

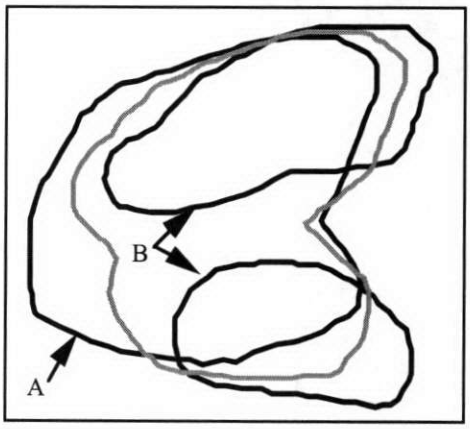


Fig. 14. Overlapping and branching contours A and B. A is at one level and B (two separate contours) at another level. The gray contour is the interpolated result.

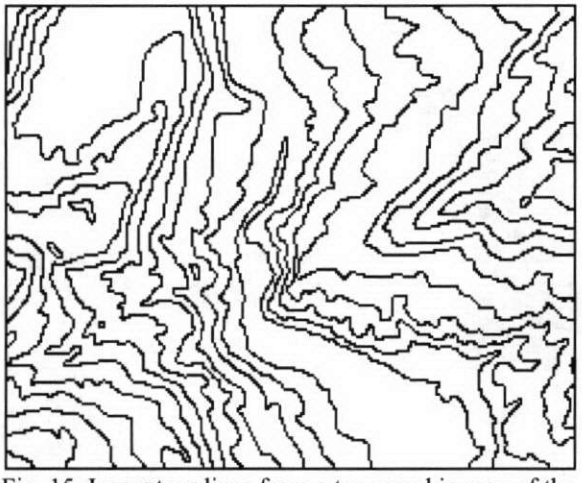


Fig. 15. Isocontour lines from a topographic map of the mouth of Provo Canyon, Utah.

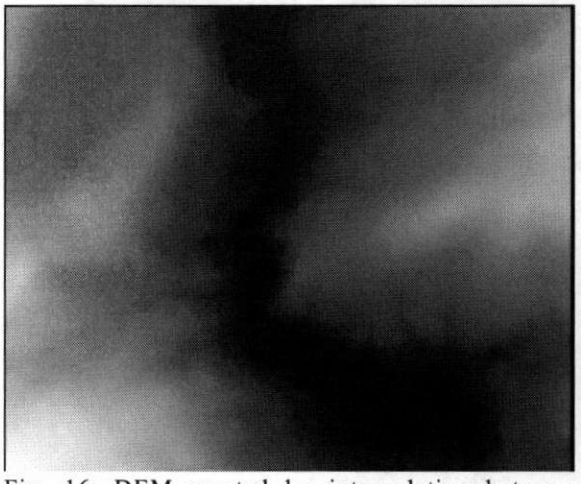


Fig. 16. DEM created by interpolating between isocontour lines in Figure 15.

As an additional illustration of robustness the algorithm was applied to digital elevation contours of Taiwan (Fig. 19). The resulting DEM (Fig. 20) preserves terrain features with striking detail.

The 500x400 DEM in Fig. 16 was created in about 2 minutes using a MATROX Image Series board hosted by a 386 processor. The 1024x1024 DEMs in Figures 18 and 20 were generated in about 5 minutes.

The algorithm has also been successfully applied to medical imaging. Figure 21 shows contours interpolated from cross-sectional outlines of the left ventricular chamber of the heart derived from Cine' CT scans. Unlike the nested isocontours used to generate terrain models, the left ventricular contours demonstrate both overlapping and branching situations. However, these present no special problems because of the general nature of the morphological interpolation as illustrated above. The resulting 3D surface of the left ventricle in Figure 22 helps to substantiate this claim.

Specific contributions of this research to applications which require generation and use of an arbitrarily large three-dimensional (3D) data base include:

1. Automated vs. manually-assisted creation of the 3D database
2. Interpolation of the height grid directly from the contour image in image space without the need to explicitly extract contour data.
3. No additional data structures necessary for intermediate representation or storage of contour data
4. It makes essential use of contour morphology.

Several computational advantages also emerge as an inherent part of the algorithm:

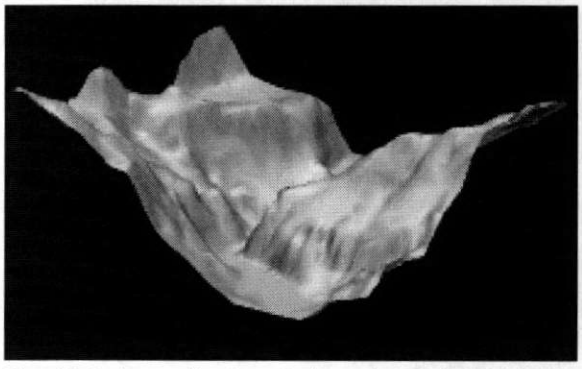


Fig. 17. Polygonalized rendering of DEM in Fig. 16.

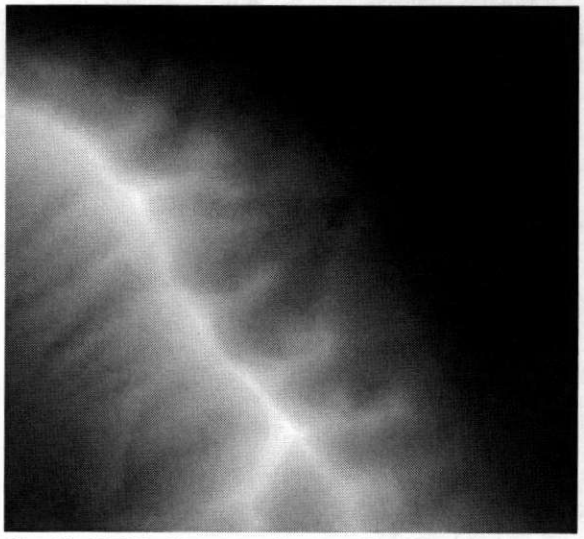


Fig. 18. DEM of Mt. Rundle (Banff, Alberta) generated by interpolating between isocontour lines.

Fig. 19. Digital isocontours of Taiwan.

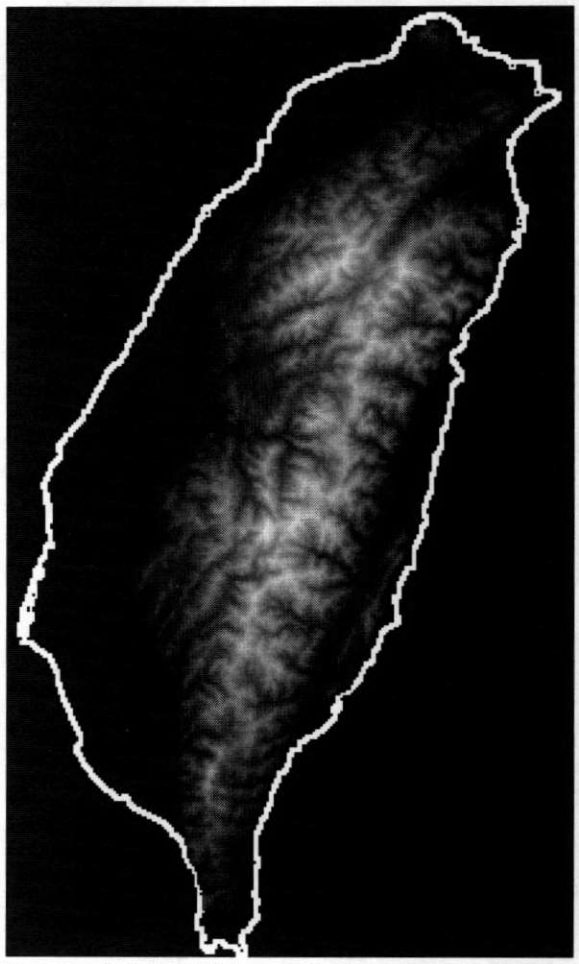


Fig. 20. DEM of Taiwan obtained by interpolating contours in Fig. 19. Coastline inserted for reference.

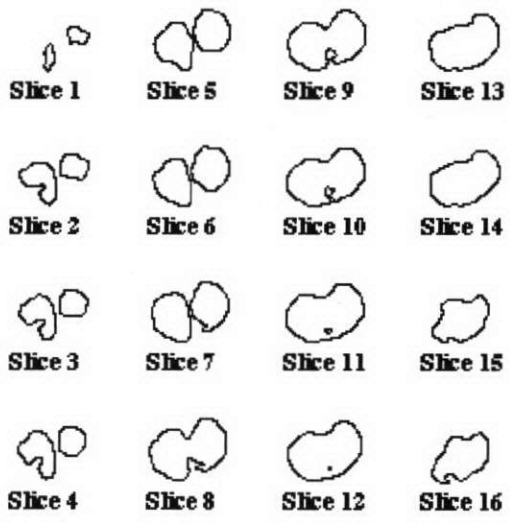


Fig. 21. Interpolated left ventricular contours.

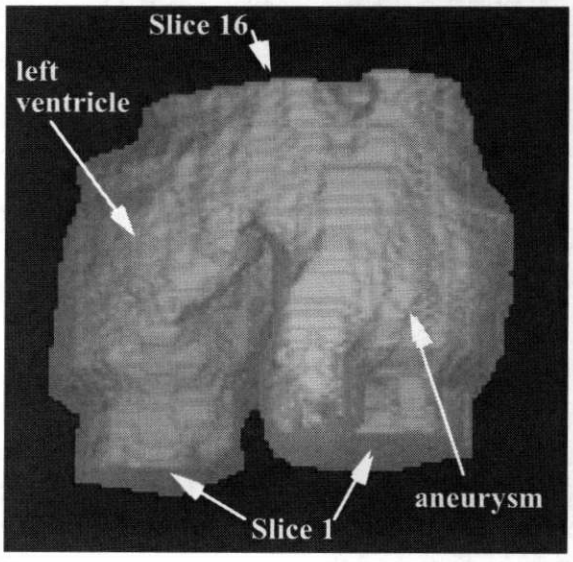


Fig. 22. 3D surface rendering of interpolated contours.

1. Contour intervals are independent of each other and can therefore be processed in parallel.
2. Recursive decomposition of each contour interval into subintervals means each contour interval can be interpolated with ~$\log_2 n$ operations (where n = width of the contour interval). This also means that the number of parallel operations increases from 1 to 2^n for each successive subinterval, at which point the contour interval is fully interpolated.
3. The algorithm is inherently parallel since for an N*N image an average of 2N points are interpolated simultaneously (ie. as a family of contour/isoline points).

5. Conclusion

Computational advantages of morphological interpolation have been presented and discussed. Other (application) advantages might include the use of contours as a highly compact/encoded representation of terrain (for an on-board data base for flight simulation). Future work will include experimentation with new/different structuring elements to distribute and balance 4-connected pixel propagation. Application to overlapping and unnested nonoverlapping contours will also be investigated. Efforts to find a transform which will generate a true (spineless) "mid"-axis transform (rather than medial axis) may be useful. One approach to producing smoother junctures (i.e. minimize contour "ghosts") would be to propagate, in a weighted sense, the gradient of the interpolated values from one contour interval into the next.

References

1. P. Yoeli. Computer executed production of a regular grid of height points from digital contours. The American Cartographer, 13(3), 219-229, 1986.
2. D. Legates and C. Willmott. Interpolation of point values from isoline maps. An American Cartographer, 13(4), 308-323, October, 1986.
3. J. Nelson, et al. Enhanced contour-to-grid interpolation procedures used in producing high-fidelity digital terrain models. Technical Papers 1988 American Congress on Surveying and Mapping - American Society for Photogrammetry and Remote Sensing Annual Convention, 1, 172-176.
4. R. Rinehart and E. Coleman. Digital elevation models produced from digital line graphs. Technical Papers 1988 American Congress on Surveying and Mapping - American Society for Photogrammetry and Remote Sensing, 2, 291- 299, 1988.
5. E. Mortensen, B. Morse, and W. Barrett. Adaptive Boundary Detection Using "Live-Wire" Two Dimensional Dynamic Programming, *IEEE Computers in Cardiology,* Durham, North Carolina, October, 1992.
6. T. Sederberg,and E. Greenwood. A Physically Based Approach to 2-D Shape Blending, SIGGRAPH Proceedings, pp. 25-34, 1992.
7. F. Verbeek, et. al. 3D Base: A Geometrical Data Base System for the Analysis and Visualisation of 3D-Shapes Obtained from Parallel Serial Sections Including Three Different Geometrical Representations, Computerized Medical Imaging and Graphics, Vol. 17, No. 3, pp. 151-164, 1993.
8. N. Kehtarnavaz, L. Simar and R. De Figueiredo. A syntactic/semantic technique for surface reconstruction from cross-sectional contours. Computer Vision, Graphics, and Image Processing, 42, 399-409, 1988.
9. S. Ganapathy and T. Dennehy. A new general triangulation method for planar contours. Computer Graphics, 16(3), 69-75, July, 1982.
10. H. Fuchs, Z. Kedem, and S. Uselton. Optimal surface reconstruction from planar contours. Communications of the Association for Computing Machinery, 20(10), 693-702, October, 1977.
11. R. Haralick and L. Shapiro. Computer and Robot Vision. Vol. I, Addison Wesley, 1992.
12. J. Serra. Image Analysis and Mathematical Morphology. Academic Press, 1982.
13. Mortensen, E. and Barrett, W. Morphological Interpolation Between Contours, *SPIE Proceedings of Medical Imaging VII,* Newport Beach, February, 1993.
14. Petersen, S., Barrett, W., and Burton, P. A New Morphological Algorithm for Automated Interpolation of Height Grids from Contour Images. SPIE/SPSE Symposium on Electronic Imaging: Science and Technology, Santa Clara, February, 1990.
15. Blum, H. Biological shape and visual science. Journal of Theoretical Biology, 38(1), 205-287, 1973.
16. Zhang, T., Suen, C. A fast parallel algorithm for thinning digital patterns. Image Processing and Computer Vision, 27(3), 326-329, March, 1974.
17. Taylor, D. and Barrett, W. An Algorithm for Continuous Resolution Polygonalizations of a Discrete Surface. GI '94 (This Proceedings.)
18. S. Raya and J. Udupa, "Shape-based Interpolation of Multidimensional Objects," *IEEE Trans. on Medical Imaging*, 9(1) :32-42, 1990.

Int. J. Radiation Oncology Biol. Phys. Vol. 21, pp. 71–78
Printed in the U.S.A.

0360-3016/91 $3.00 + .00

● *Original Contribution*

DOSE-VOLUME HISTOGRAMS

R. E. DRZYMALA, PH.D.,[1] R. MOHAN, PH.D.,[2] L. BREWSTER, M.S.,[2] J. CHU, PH.D.,[3] M. GOITEIN, PH.D.,[4] W. HARMS, B.S.[1] AND M. URIE, PH.D.[4]

[1]Mallinckrodt Institute of Radiology, Washington University School of Medicine, St. Louis, MO 63110; [2]Memorial Sloan-Kettering Cancer Center, New York, NY 10021; [3]University of Pennsylvania School of Medicine and the Fox Chase Cancer Center, Philadelphia, PA 19111; and [4]Massachusetts General Hospital, Department of Radiation Medicine, Boston, MA 02114 and Harvard Medical School

A plot of a cumulative dose-volume frequency distribution, commonly known as a dose-volume histogram (DVH), graphically summarizes the simulated radiation distribution within a volume of interest of a patient which would result from a proposed radiation treatment plan. DVHs show promise as tools for comparing rival treatment plans for a specific patient by clearly presenting the uniformity of dose in the target volume and any hot spots in adjacent normal organs or tissues. However, because of the loss of positional information in the volume(s) under consideration, it should not be the sole criterion for plan evaluation. DVHs can also be used as input data to estimate tumor control probability (TCP) and normal tissue complication probability (NTCP). The sensitivity of TCP and NTCP calculations to small changes in the DVH shape points to the need for an accurate method for computing DVHs. We present a discussion of the methodology for generating and plotting the DVHs, some caveats, limitations on their use and the general experience of four hospitals using DVHs.

Dose-volume histograms, Radiation therapy, Computerized treatment planning.

INTRODUCTION

Three-dimensional treatment planning involves a large body of dose information made available by volume dose calculations. The sheer volume of information may make it difficult to interpret and assimilate the data displayed as isodoses on a number of transverse, sagittal, coronal or oblique planes, as three-dimensional isodose surfaces, or as other forms of three-dimensional displays. Condensing the three-dimensional dose distribution data into dose-volume histograms enables one to graphically summarize the radiation distribution throughout the target volume and the anatomical structures of interest. A histogram may be plotted according to the usual mathematical definition, as the accumulated volume of those elements receiving dose in a specified dose interval against a set of equispaced dose intervals. This is referred to as a differential dose-volume histogram. We have found it more useful, however, to plot the data as the volume receiving a dose greater than or equal to a given dose against that dose over the expected dose range. These plots are actually cumulative dose-volume frequency distributions; we shall hereafter refer to them simply as dose-volume histograms (DVHs). In most instances, the volume is specified as the percentage of the total volume of a structure receiving dose within each interval; however, it may be more appropriate to specify absolute volume in some cases.

DVHs may be valuable in the planning process to check if dose is at an adequate level and uniform throughout the target volume, or for revealing the presence and extent of hot spots in adjacent normal tissues. DVHs may be used as a preliminary step in evaluating a treatment plan, or as a screening tool to select the best or most acceptable plan(s) from a group of plans before more extensive information is examined in the form of 2-D or 3-D isodose displays. [They may also be used as a graphical way of comparing different treatment plans in a single plot, and to produce measures of tumor control probability (TCP) or normal tissue complication probability (NTCP) (10, 11), allowing quantitative scoring and evaluation of plans.]

The use of dose-volume histograms has been reported in the literature (1, 3, 4, 6, 8, 10–12, 15) for the pancreas and prostate, but details of methods for their calculation have not been described. Four hospitals have participated in an NCI-sponsored contract studying three-dimensional photon treatment planning, and we present here their DVH techniques and experiences.

Supported in part by NCI Contracts NO1 CM-47316, NO1 CM-47695, NO1 CM-47696, NO1 CM-47697.

Reprint requests to: R. E. Drzymala, Mallinckrodt Institute of Radiology, Washington University School of Medicine, 510 S. Kingshighway Blvd., St. Louis, MO 63110.

METHODS AND MATERIALS

Each of the four institutions involved in this study developed their histogram code independently. Most of the features implemented are shared by all institutions, with differences occurring primarily in the handling of boundary regions and resolution. DVHs are calculated as follows: first, the boundaries of pertinent anatomical structures of the patient must be defined and the dose must be computed for the volume of interest. The anatomy is then subdivided into a volume grid of appropriate resolution. To arrive at the total volume for a particular structure, one sums those volume elements (voxels) lying within the structure. The dose for each voxel can be determined concurrently with voxel-summing so that, ultimately, all contributing voxels are accumulated within the appropriate dose bin of the histogram for each structure. The bin values are then plotted.

Anatomical representation

The three-dimensional image, usually composed of a contiguous sequence of computed tomographic (CT) scans, provides the basic anatomical representation of the patient. One can think of this volume as being subdivided into a three-dimensional grid of cuboid volume elements which is equispaced in the X and Y directions but may not be equispaced in the Z direction. (In our convention, each CT slice lies at a constant Z and each element in the slice is referenced by the X and Y coordinates of a point at its center.) One may divide the volume into contiguous "slabs" formed by the CT slices. These form a natural set of planes since it is through the CT images that the structures of interest are defined. If the scans are considered to just fill the volume they span, then the thickness of the n'th slab is equal to the distance between the $(n-1)$th and the $(n+1)$th slices. (The first and last scans have a thickness equal to their spacing from their nearest neighbor.) The volume of a voxel, therefore, is the product of the X and Y grid spacings and the slab thickness.

Optimum grid spacing and size

Computing a dose-volume histogram involves three-dimensional anatomical volume and radiation dose matrices. The three-dimensional matrix used for volume calculations need not be coincident, however, with the three-dimensional dose matrix nor the three-dimensional matrix of CT voxels. Dose calculations tend to be slow; usually a relatively coarse matrix is employed, but the matrix required for accurate volume estimation may be of much finer resolution. The accuracy of volume calculation, therefore, should not be limited by the dose matrix resolution. One can estimate the dose to each voxel of appropriate resolution by tri-linear interpolation from the dose matrix to the center of the voxel, thereby avoiding the need for a dose matrix with resolution equal to the volume grid. Dose resolution can then depend solely upon dose gradients in the region. If the voxel lies within a structure but not within the dose matrix, i.e., outside the calculation window, the dose to that voxel is considered to be zero. This can be a problem if the planner does not cover the entire region of interest receiving significant dose with the dose calculation window.

Volume. Attempts were made to determine the volume resolution necessary to achieve an acceptable level of accuracy through analyzing DVHs for various structures and manipulating the voxel size (4). As would be expected, small structures appeared most sensitive to increases in voxel size. Experimental data indicate that, for a constant dose matrix of comparable resolution, little change is evident in the DVH for the spinal cord residing in a fairly homogeneous dose region for a Hodgkins treatment plan if the voxel dimensions are less than 0.25 that of the structure. Volume estimation was on the average within 5% of the true value, but theory predicts errors up to 16% for extreme cases at this resolution. This results primarily from the errors accumulated while detecting the location of irregular contour boundaries with a regular rectangular grid of voxels. Changes of structure position in the volume grid help, since accumulated errors in boundary estimation average out. Accuracy, however, depends also upon the minimum of the structure, where the presence of a long, narrow but important appendage may require a finer resolution. Since greater resolution means a smaller voxel size and increased computation time (particulary in large structures), one should, as one group did, dynamically adjust the voxel size to the dimensions of the structure independently in X, Y and Z. This may be accomplished through the use of quadtrees and/or octrees.

Quite often the CT scan series covers only a portion of the organ of interest, which, if unaccounted for, can be a problem when interpreting DVHs. One possible solution involves the use of a set of standard volumes against which a computed volume could be compared. If the computed volume is significantly smaller than the standard size for the organ, additional volume can be added to the zero dose bin. This assumes, of course, that the CT series covers the region getting any significant dose.

Dose. The selection of grid spacing for the three-dimensional dose matrices involves compromise. The finer the grid, the greater the accuracy of dose calculation in regions of high dose gradients. The disadvantages of fine grid spacing include increased computation times and the creation of large disk files. In general, one must specify two criteria: 1) the desired dose accuracy, and 2) the maximum acceptable distance between the estimated and actual isodose contours. A dose interpolation would be acceptable if either criterion were met. The most stringent situation clearly occurs in regions of rapidly varying dose, for example, at the beam margins. A calculation of the error in the dose estimate when interpolating within regular rectangular grid indicates (Niemierko and Goitein, private communciation) that the required grid size depends on the detailed shape of the beam profile. In both the steepest and

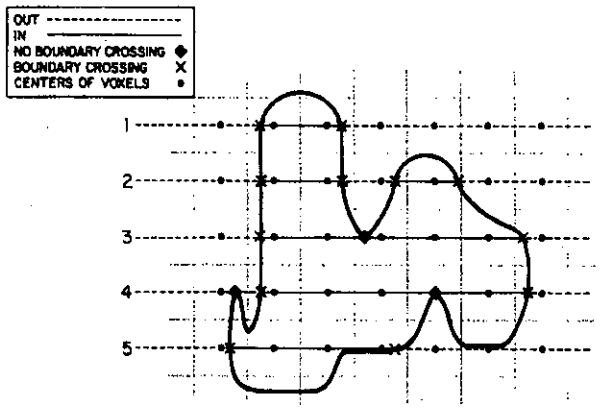


Fig. 1. Diagram showing scan lines (red) traversing an irregularly shaped contour (black) and a volume grid (blue) superposed. Each scan line demonstrates a different situation that the sampling algorithm must handle while searching for contour boundary crossings. Of particular importance are the conditions in lines 3, 4 and 5 where not all intersections of the scan line with the contour are boundary crossings.

most slowly varying regions, linear interpolation between grid points is a good approximation. It is in the intermediate regions, where the second gradient of the dose profile is large, that the greatest error is made. The best choice of grid spacing to achieve a given accuracy is dominated by the greatest allowable isodose position error, rather than the accuracy of dose estimation at a point. For most practical beams, in order to achieve an isodose positioning error of $\leq$ to delta—or a dose accuracy of $\leq$ to 2%, a grid spacing of about 2.5 times delta is needed. (Provided delta is $\geq$ 2 mm.) Thus 2% dose accuracy or 2 mm isodose positional accuracy (but not necessarily both) will be achieved with a grid spacing of 5 mm.

Assignment of voxels to a structure

Contours drawn on a contiguous parallel set of CT images usually define the structures. The cartesian volume matrix used to calculate the DVH is defined so that its Z-axis is normal to the sections on which the structure is defined. To determine which points of the volume matrix are within a structure, we consider one (X-Y plane) out of the volume matrix at a time and looping over each row equal to the y value at the center of the volume matrix voxels, find the intersections of the row with the contour(s) outlining the structure. If the X-Y planes of the volume matrix are not coincident with the CT slices, the outline(s) of the structure is obtained by interpolation between the closest CT sections on which the structure is defined. The algorithm selects and sorts only the transections (crossings of the boundary by the row) in the ascending order of distance from the beginning of the row. The voxels on the row whose centers lie between the odd numbered and the next even numbered transection are within the current structure while those whose centers follow the even numbered transections are outside the structure (see Fig. 1).

The volume of elements within a structure may accumulate in the appropriate dose interval bins "on the fly" as their location within a structure is ascertained. Alternatively, a three-dimensional matrix of tags identifying the structure (or segment of a multi-segment structure) of each volume element may be created prior to DVH calculations (12). The latter scheme is useful when combinations of structures are to be considered.

Binning

The range of expected dose values is divided into equispaced intervals in constructing a histogram. For each interval, the volumes of the voxels receiving dose within that interval are accumulated in the appropriate element of an array or bin. At the completion of the calculation, each bin in the array contains the summed volume of the voxels which received dose within the corresponding interval. Cumulative DVHs are obtained by adding volumes accumulated in each bin with the volumes in all bins corresponding to higher dose intervals, whereas to obtain differential DVHs, the volumes accumulated in the bins are not added.

For cumulative DVHs, an appropriate dose interval for the bins depends on the dose response curve for the structure of interest. It appears that 0.5 Gy ($\sim$ 1% of the prescription dose) is reasonable, whereas 2 Gy is crude. The group which used differential dose-volume histograms extensively concluded that, for this type of histogram, a crude bin interval of 2 to 5 Gy was advantageous in evaluating our breast plans.

Errors in calculating DVHs

The accuracy with which a DVH is estimated depends on the accuracy of estimating the boundaries of structures, the accuracy of dose calculations and interpolations of the dose matrix. Subjectivity and accuracy with which we outline structures on CT scans must all be considered. Where structure boundaries are not readily apparent, the process of structure outlining may be quite unreproducible. Contours are often entered into the computer with hard-to-control input devices, such as a mouse, joystick, trackball or digitizer pad, which introduces additional error. Improvements, including computer automation in some situations, are needed.

Since structures of interest can be quite irregularly shaped, the algorithm used to detect transections of the scan line with the structure contours is crucial. Not all intersections are transections. For example, invaginations and evaginations of the contour that touch the scan line then reverse direction should not be considered transections (Fig. 1, line 3). Also regions where the contour and scan line proceed superimposed (Fig. 1, line 5) are not transections except at one point where the contour traverses the scan line. Although the latter might seem like a rare special case, it does occur because of the discreet representation of structures on a pixel-based image.

A direct way to verify the DVH computation code before implementation is to create a few test phantoms of

known dimensions and dose assignments. A set of nested, concentric cubes was useful as a phantom; each rind formed therein had a different uniform dose. With the sides of the cubes oriented in parallel or perpendicularly to the scan line, this phantom provides a rigorous test for registration of the dose and volume matrices with the structure contours. This is particularly true if the inner cubes are excluded from the outer during the DVH calculation. Another appropriate test phantom is a cube with a rectangular notch cut in at one side. It is sensitive to boundary detection and the algorithm's ability to discriminate transections from intersections if a scan line is forced to coincide with the bottom edge of the notch. To complete the test, a cylindrical test phantom is helpful to assure the accuracy of the calculation for noncuboid shapes.

Addition and subtraction of DVHs

In some cases, volumes which are defined by some combination of more than one defined structure are of interest. For example, a DVH combining both right and left lungs may be desirable. In the case of overlapping structures, flexibility in structure definition is helpful. For instance, contours defining a patient's rectum and those defining a rectal tumor may overlap at some CT levels. One might want a histogram of all those points within the rectum, or only those points within the rectum but outside the tumor volume. On the other hand, a union of both structures may be of interest. Similar situations can occur when one structure is completely contained within another.

Separation and combination of structures could occur at the time of drawing their contours on CT images, but often this is not a practical nor the most flexible approach. One method of specifying and computing DVHs of such combinations of structure is by using a system of logical operators, such as 'and,' 'or,' 'not,' 'inside,' and 'outside.' Using the example of rectal cancer, the points within the rectum but not within the target would be specified as 'inside rectum and outside target.' The sum of both structures would be 'inside rectum or inside target.' The process of DVH calculation then becomes a matter of determining, for each point of the volume grid, whether the point lies within the logically specified region.

One method has been developed which allows the isolation of individual structures, as well as addition and subtraction operations on structures. It involves the construction of a user-specified hierarchy among anatomical structures prior to histogram calculation. The hierarchy level assigned to each structure is used to determine in which structure a given dose matrix point is considered to rest. This is necessary when structures overlap at the point.

The user assigns a hierarchy number to each anatomical structure of interest, with a higher number corresponding to a higher priority. This is done interactively for each series of histograms desired. To assign each grid point in the three-dimensional dose matrix to the structure containing the point, a corresponding matrix of structures, "tag," is created (12). Each element of the tag matrix has the same position in space as the corresponding element in the dose matrix, and is given a value related to an anatomical structure and its segment number. Since the tag matrix points are assigned values in ascending order of hierarchy, points common to more than one region will be set to the tag value corresponding to the structure with the highest hierarchy. This procedure results in a 3-D matrix of tags associating a structure with each dose point, thereby telling us in which structure any given dose matrix point is considered to rest.

With ambiguities thus resolved, the addition of two disjointed structures is accomplished by simply adding the corresponding dose level bins for the two structures, while the addition and subtraction of overlapping structures is done by proper assignment of hierarchies, and is carried out in the production of the structure tag matrix. For example, a structure contained within another can be subtracted out by assigning the inner structure a higher priority.

RESULTS

Comparisons among the institutions

Three of the four institutions participated in a comparison of the different calculation methods for producing cumulative DVHs. Each participant used the same concentric cube phantom containing contours and doses as described above. The resulting plots were essentially the same ($\pm$ 1 bin), with differences being attributable to the different resolutions employed and the way in which each program handled the boundaries of structure contours. It is difficult to assess the importance of the differences to clinical evaluation of the plans. We have observed, however, that small differences between histograms can affect TCP and NTCP calculations significantly (see also Discussion and Conclusion).

Examples of cumulative dose-volume frequency distributions

Figures 2 through 5 are examples of cumulative dose-volume histograms. We should mention some useful aspects of the graphic presentation. We have found it convenient to include both absolute dose and volume along with normalized relative values on the same plot. This can be easily accomplished by labeling four sides of the plot. An overlaid grid (not shown) helps to make the inspection of the curves more quantitative. When comparing different treatment plans, it is advantageous and, indeed, necessary to superimpose the DVH curves, labeled and appropriately color coded and/or patterned, on the same plot (Fig. 5). Differences are much more obvious when overlaid on the same plot. It is no accident that the one group which did not do this found DVHs less helpful in their assessment than did the other three groups.

Figure 2 is an example of a DVH computed for a target

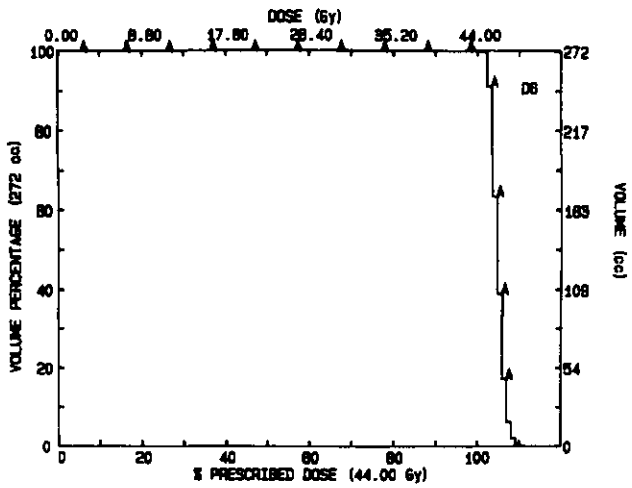


Fig. 2. An example of a cumulative dose-volume frequency distribution or histogram (DVH) computed and plotted for a structure that has a uniformly high dose throughout its volume. Typically, the volume is summed with that in intervals of higher dose and plotted against the dose intervals. The volume and dose may be absolute or normalized values, normalized to total volume or prescribed dose, respectively.

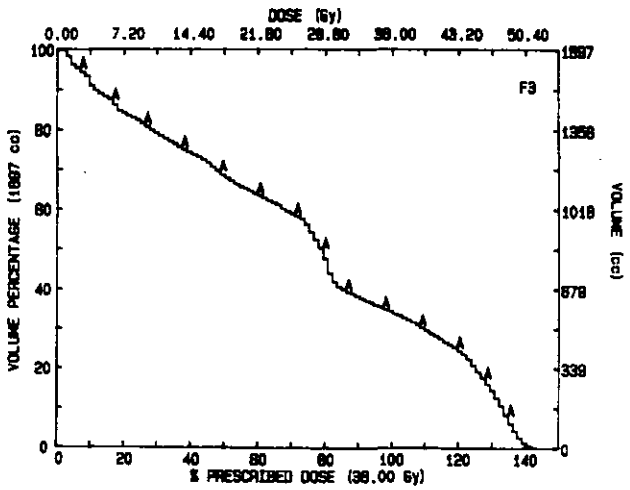


Fig. 3. The DVH depicted in this figure represents a structure having a dose distribution that is heterogeneous over the dose range. This is exemplified by a slope of approximately 45 degrees over much of the dose range. The plot makes no statement about the location of the dose. It could result from the structure residing in a dose gradient region or a dotted distribution of variable dose. (This is, however, a rather unlikely occurrence for teletherapy photon beams.)

volume which has a high, fairly uniform dose. The DVH under such circumstances approximates a step-function. Large, steep drops in the DVH ordinate are indicative of a large percentage of the volume having a similar dose. A plan giving this shape DVH for the target volume with the step at or just above the prescription dose would be close to ideal. This is contrasted with the curve of Figure 3 where we have a relatively shallow and constant slope. This curve represents a heterogeneous dose distribution in the volume of interest. The concave curve in Figure 4 represents a structure, the majority of which receives a low dose, but with one or more small hot spot(s). Not shown is the trivial case where the structure receives little or no dose. This case would appear as a step from 0 to 100% in the left-most bin and verifies that the structure is far removed from the radiation beams and their scattered radiation.

Differential dose-volume histograms (not shown) are true histograms. These plot the absolute or relative volume in each dose interval directly which can facilitate comparisons between dose bins of a histogram. True histograms also have a finer structure, possibly offering more detail for analysis. When multiple histograms from different plans are overlaid, the appearance of differential histograms is confusing. This is less of a problem with cumulative histograms. Plotting the cumulative dose-volume histogram as a "smoothed" line graph in contrast to its choppy appearance in the figures would serve to distinguish it further from the plots of true histograms. One approach would be to connect the right-most corners of adjacent bins with a straight line. One should remember that the plots should reflect the units of volume per dose interval, not discreet dose values. Using a nonlinear volume axis when plotting cumulative dose-volume histograms may help to observe fine structure in some regions of the curves (see Fig. 6).

Limitations of dose-volume histograms and some alternatives

DVHs graphically summarize dose distribution within an anatomic structure. While they provide information on the existence and magnitudes of hot and cold spots, they indicate neither where, within a structure, a hot spot

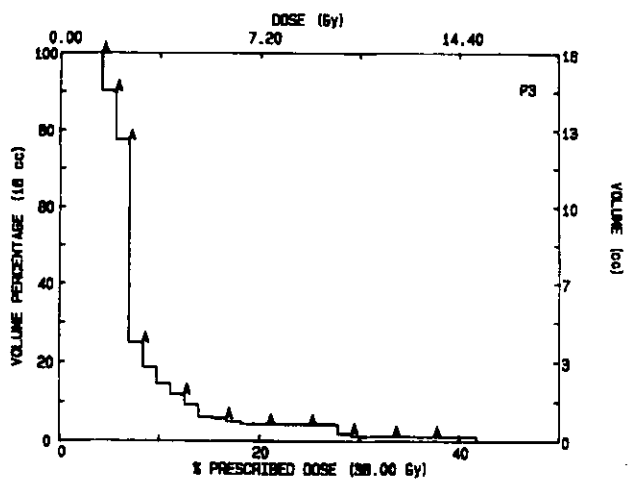


Fig. 4. The concave appearance of this DVH is indicative of a structure receiving predominantly low dose, but with one or more relatively small, higher dose regions.

 3770

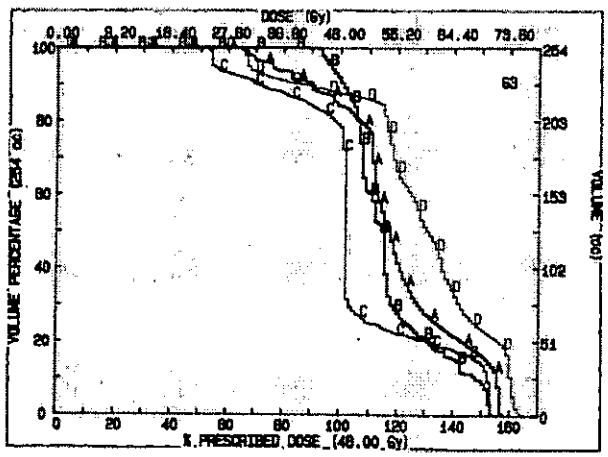


Fig. 5. Superimposed DVHs for the bladder comparing different treatment plans for cancer of the prostate. Color coding and repetitive labeling distinguish the curves.

occurred, nor whether a hot spot occurred in one or several disconnected regions. This may be more of a problem for some organs than others. The location of hot and cold spots in a structure may be dissected with other approaches.

In the spinal cord, for instance, a hot spot which transects the cord could have very different implications than one which was purely superficial. A solution to this problem might be a plot of the radial extent of the dose above a critical threshold as a function of length along the cord (again with a reduction in dimensionality).

The Dose Area Frequency Distribution (or Dose Area Histogram) is appropriate for skin and other anatomic surfaces, such as the rectum and bladder. DAHs are similar to DVHs, but with a reduction in dimensionality. They could be used in the planning process to predict and evaluate skin reactions for various types of treatments. In this case, the histogram would accumulate the surface areas of elements of the skin in the area of treatment as a function of the dose they received. As with volumes, contours can be used to define the surface of interest. The dimensions of each surface element are obtained from the spacing along the contour and between contours (i.e., the gaps between CT slices), with the curvature of the body, along the contour line as well as perpendicular to it, necessarily taken into account. The accurate estimation of dose at points on the sloping skin surface (such as in tangential breast treatments) is a complex process.

Dose-length histograms may be of value for long, linear structures such as the optic nerves or their blood vessels.

Another limitation of DVHs is that, as with dose distributions, interpretation of the plot is rather subjective and the implications of small differences between DVHs are poorly understood at present. This would point to the value of an objective numeric score, such as tumor control probability or normal tissue complication probability, to provide objective rank when comparing plans. Furthermore, we have found that small changes in the DVHs at times resulted in significant differences in the computed TCP and NTCP values.

DVHs do not reflect the complexity of the field arrange-

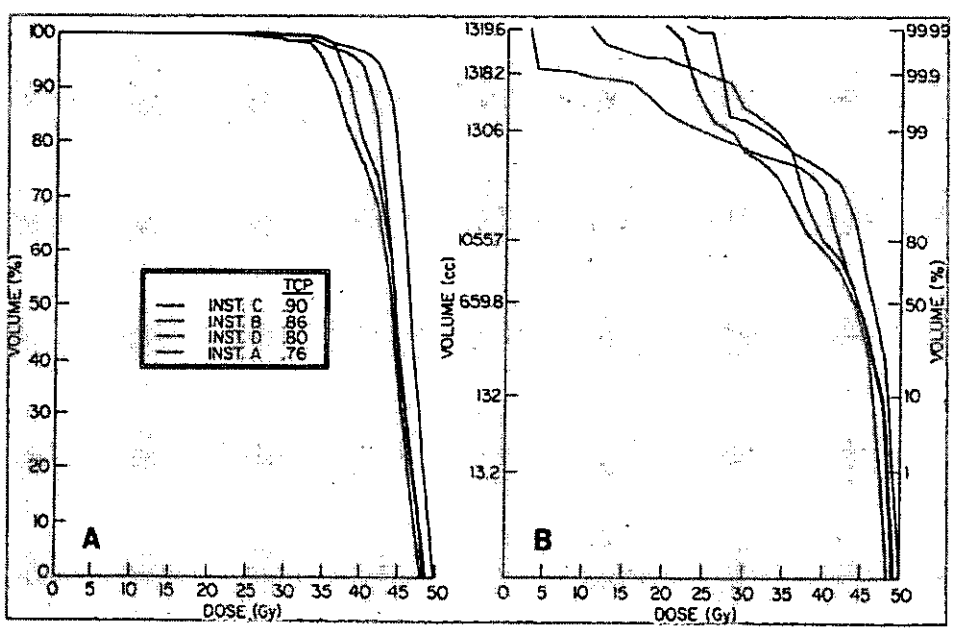


Fig. 6. In panel A are typical cumulative dose-volume histograms as normally plotted and superposed for use in this study. The inset gives the corresponding TCP calculations for the curves. Inspection of the curves as plotted does not reveal a good correlation of the curve positions with these numbers (see text). Replotting the volume axis with an expanded (panel B) scale shows a correlation with the small percentage of volume elements in the lower dose ranges.

 3771

ments, and, unlike dose distribution, this complexity cannot be inferred from the DVHs.

DISCUSSION AND CONCLUSION

The value of Cumulative Dose-Volume Frequency Distributions, less precisely termed Dose-Volume Histograms, has been investigated at four institutions participating in an NCI-sponsored contract exploring three-dimensional photon treatment planning. The usefulness of DVHs has been evaluated for eight anatomic sites. The specific findings for these and other sites are described elsewhere (1, 2, 3, 5, 7–9, 13–17). In our opinion, DVHs are a valuable means of summarizing, in graphic form, the large body of dose distribution information provided in 3-D planning. Their greatest strength is their ability to provide rapid screening of plans.

The detailed implementation of DVH computations could affect the shape of the plots significantly and accurate calculations are necessary if DVHs are to be widely used in the radiation therapy community. This would impact on estimates of TCP and NTCP as well, since the DVHs provide input data for these calculations and small differences in the DVHs have been observed in some circumstances to result in significant changes in computed TCP and NTCP values. For example, comparing TCP calculations with dose-volume histograms plotted in Figure 6, it is demonstrated that doses in a very small percentage of the total volume can have a significant impact in "ranking" treatment plans by TCP values. Inspecting the dose-volume histograms (panel A) as typically plotted, one would expect that the plans should be ranked C > A > B > D. Accordingly, more of the tumor volume would receive higher doses. The TCP calculation gives the order C > B > D > A. Replotting the same data with an expanded scale for the ordinate in panel B emphasizes volumes with higher and lower doses relative to those in the middle, and demonstrates a better correlation of the curves in the lower dose region with the calculated TCPs. The significance of this is not understood; however, it indicates that we should use methods that reduce the histogram to a single number cautiously.

The lack of information regarding the spatiality of dose distribution is one major limitation of DVHs; two- and three-dimensional dose displays are required for such analysis. This fact signifies that *DVHs cannot be the sole criterion for evaluating a plan, or disclosing the best treatment plan.* Complexity of field arrangement and reproducibility of patient set up are other important factors not taken into account by DVHs. However, these factors being equal, DVHs can be used effectively to evaluate and compare rival treatment plans.

REFERENCES

1. Austin-Seymour, M. M.; Chen, G. T. Y.; Castro, J. R.; Saunders, W. M.; Pitluck, S.; Woodruff, K. H.; Kessler, M. Dose volume histogram analysis of liver radiation tolerance. Int. J. Radiat. Oncol. Biol. Phys. 12:31–35; 1986.
2. Brown, A. P.; Urie, M. M.; Barest, G.; Cheng, E.; Coia, L.; Emami, B. N.; Galvin, J.; Kutcher, J.; Manolis, J.; Wong, J. W.; Yahalom, J. Three-dimensional treatment planning for Hodgkin's disease. Int. J. Radiat. Oncol. Biol. Phys. 21: 205–215; 1991.
3. Chen, G. T. Y.; Austin-Seymour, M.; Castro, J. R.; Collier, J. M.; Lyman, J. T.; Pitluck, S.; Saunders, W. M.; Zink, S. R. Dose volume histograms in treatment planning evaluation of carcinoma of the pancreas. In: Proceedings, Eight International Conference on Uses of Computers in Radiation Therapy, IEEE, ISBN 0-8186-0559-6; 1984:264–268.
4. Chu, J. C. H.; Richter, M. P.; Sontag, M. R.; Larsen, R. D.; Fong, K.; Bloch, P. Practice of 3-dimensional treatment planning at the Fox Chase Cancer Center, University of Pennsylvania. Radiother. Oncol. 8:137–143; 1987.
5. Coia, L. R.; Galvin, J.; Sontag, M.; Blitzer, P.; Brenner, H.; Cheng, E.; Doppke, K.; Harms, W.; Hunt, M.; Mohan, R.; Munzenrider, J.; Simpson, J. Three-dimensional photon treatment planning in carcinoma of the larynx. Int. J. Radiat. Oncol. Biol. Phys. 21:183–192; 1991.
6. Drzymala, R. E.; Harms, W. B.; Purdy, J. A. Dose-volume histograms for 3-D radiation treatment plans. Med. Phys. 14(3):460; 1987 (abstr.).
7. Emami, B.; Purdy, J. A.; Manolis, J.; Barest, G.; Cheng, E.; Coia, L.; Doppke, K.; Galvin, J.; LoSasso, T.; Matthews, J.; Munzenrider, J.; Shank, B. Three-dimensional treatment planning for lung cancer. Int. J. Radiat. Oncol. Biol. Phys. 21:217–227; 1991.
8. Goitein, M.; Miller, T. Planning proton therapy of the eye. Med. Phys. 10:275–283; 1983.
9. Kutcher, G. J.; Fuks, Z.; Brenner, H.; Brown, A. P.; Burman, C.; Cheng, E.; Coia, L.; Krippner, K.; Manolis, J. M.; Mohan, R.; Simpson, J. R.; Urie, M.; Vikram, B.; Wallace, R. Three-dimensional photon treatment planning for carcinoma of the nasopharynx. Int. J. Radiat. Oncol. Biol. Phys. 21:164–182; 1991.
10. Lyman, J. T. Complication probability as assessed from dose-volume histograms. Radiat. Res. 104:s-13–s-19; 1985.
11. Lyman, J. T.; Wolbarst, A. B. Optimization of radiation therapy, III: A method of assessive complication probabilities from dose-volume histograms. Int. J. Radiat. Oncol. Biol. Phys. 13:103–109; 1987.
12. Mohan, R.; Brewster, L. J.; Barest, G. D. A technique for computing dose volume histograms for structure combinations. Med. Phys. 14(6):1048–1052; 1987.
13. Munzenrider, J. E.; Doppke, K. P.; Brown, A. P.; Burman, C.; Cheng, E.; Chu, J.; Chui, C.; Drzymala, R. E.; Goitein, M.; Manolis, J. M.; Nori, D.; Simpson, J. R.; Solin, L.; Urie, M. M. Three-dimensional treatment planning for paraaortic node irradiation in patients with cervical cancer. Int. J. Radiat. Oncol. Biol. Phys. 21:229–242; 1991.
14. Shank, B.; LoSasso, T.; Brewster, L.; Burman, C.; Cheng, E.; Chu, J. C. H.; Drzymala, R. E.; Manolis, J.; Pilepich, M. V.; Solin, L. J.; Tepper, J. E.; Urie, M. M. Three-dimensional treatment planning for post-operative treatment of rectal carcinoma. Int. J. Radiat. Oncol. Biol. Phys. 21:253–265; 1991.
15. Shipley, W. U.; Tepper, J. E.; Prout, G. R.; Verhey, L. J.; Mendiondo, O. A.; Goitein, M.; Koehler, A. M.; Suite, H. D. Proton radiation as boost therapy for localized prostatic

 3772

carcinoma. JAMA 241:1912–1915; 1979.
16. Simpson, J. R.; Purdy, J. A.; Manolis, J. M.; Pilepich, M. V.; Burman, C.; Forman, J.; Fuks, Z.; Cheng, E.; Chu, J.; Matthews, J.; Mohan, R.; Solin, L.; Tepper, J.; Urie, M. Three-dimensional treatment planning considerations for prostate cancer. Int. J. Radiat. Oncol. Biol. Phys. 21:243:–252; 1991.
17. Solin, L. J.; Chu, J. C. H.; Sontag, M. R.; Brewster, L.; Cheng, E.; Doppke, K.; Drzymala, R. E.; Hunt, R. E.; Kuske, R.; Manolis, J. M.; McCormick, B.; Munzenrider, J. E. Three-dimensional treatment planning of the intact breast. Int. J. Radiat. Oncol. Biol. Phys. 21:193–203; 1991.

 3773

Food and Drug Administration
Office of Device Evaluation &
Office of In Vitro Diagnostics

COVER SHEET MEMORANDUM

From: Reviewer Name John Chen

Subject: 510(k) Number K102915

To: The Record

Please list CTS decision code _____

- ☐ Refused to accept (Note: this is considered the first review cycle, See Screening Checklist http://eroom.fda.gov/eRoomReg/Files/CDRH3/CDRHPremarketNotification510kProgram/0_5631/Screening%20Checklist%207%202%2007.doc)
- ☐ Hold (Additional Information or Telephone Hold).
- ☐ Final Decision (SE, SE with Limitations, NSE, Withdrawn, etc.).

Please complete the following for a final clearance decision (i.e., SE, SE with Limitations, etc.):		YES	NO
Indications for Use Page	*Attach IFU*	✓	
510(k) Summary /510(k) Statement	*Attach Summary*	✓	
Truthful and Accurate Statement.	*Must be present for a Final Decision*	✓	
Is the device Class III?			X
If yes, does firm include Class III Summary?	*Must be present for a Final Decision*		
Does firm reference standards? (If yes, please attach form from http://www.fda.gov/opacom/morechoices/fdaforms/FDA-3654.pdf)			X
Is this a combination product? (Please specify category _____, see http://eroom.fda.gov/eRoomReg/Files/CDRH3/CDRHPremarketNotification510kProgram/0_413b/COMBINATION%20PRODUCT%20ALGORITHM%20(REVISED%203-12-03).DOC			
Is this a reprocessed single use device? (Guidance for Industry and FDA Staff – MDUFMA - Validation Data in 510(k)s for Reprocessed Single-Use Medical Devices, http://www.fda.gov/cdrh/ode/guidance/1216.html)			
Is this device intended for pediatric use only?			
Is this a prescription device? (If both prescription & OTC, check both boxes.)		✓	
Did the application include a completed FORM FDA 3674, Certification with Requirements of ClinicalTrials.gov Data Bank?			
Is clinical data necessary to support the review of this 510(k)? Did the application include a completed FORM FDA 3674, *Certification with Requirements of ClinicalTrials.gov Data Bank*? (If not, then applicant must be contacted to obtain completed form.)			
Does this device include an Animal Tissue Source?			
All Pediatric Patients age<=21			
Neonate/Newborn (Birth to 28 days)			
Infant (29 days -< 2 years old)			
Child (2 years -< 12 years old)			
Adolescent (12 years -< 18 years old)			
Transitional Adolescent A (18 - <21 years old) Special considerations are being given to this group, different from adults age ≥ 21 (different device design or testing, different protocol procedures, etc.)			

Rev. 7/2/07

Transitional Adolescent B (18 -<= 21; No special considerations compared to adults => 21 years old) ✓

Nanotechnology

Is this device subject to the Tracking Regulation? (Medical Device Tracking Guidance, http://www.fda.gov/cdrh/comp/guidance/169.html) *Contact OC.* ✓

Regulation Number	Class*	Product Code
892.5750	II	MUJ

(*If unclassified, see 510(k) Staff)

Additional Product Codes:

Review: _____ (Branch Chief) DRAD (Branch Code) 1/12/11 (Date)

Final Review: _____ (Division Director) 1/12/11 (Date)

K102915 Premarket Notification [510(k)] Review

Date: 12/03/10

To: The Record

Office: ODE

From: John Chen, Ph. D.

Division: OVID/DRAB

510(k) Holder: ViewRay, Inc.

Device Name: ViewRay Treatment Planning and Delivery System

Contact: Janice brownlee Phone: 440-703-3210 Fax: 440-703-3229

Email: jkbrownlee@viewray.com

I. Purpose and Submission Summary

The 510(k) holder would like to introduce ViewRay Treatment Planning and Delivery System into interstate commerce.

II. Administrative Requirements

	Yes	No	N/A
Indications for Use page (Indicate if: Prescription or OTC)	X		
Truthful and Accuracy Statement	X		
510(k) Summary or 510(k) Statement	X		
Standards Form	X		

III. Device Description

The ViewRay Treatment Planning and Delivery System (TPDS) provide tools for planning and delivery of external gamma beam stereotactic radiosurgery and precision radiotherapy for lesions, tumors, and conditions anywhere in the body when radiation treatment is indicated. The ViewRay TPDS is capable of assisting clinicians in reviewing, prescribing, tracking, and correcting the course of patient treatment using tools for contouring, visualization, data storage, anatomical target monitoring and re-optimization.

The ViewRay System is an MR Image□guided radiation therapy device. Similar to other existing image - guided radiation therapy (IGRT) systems in the marketplace, the ViewRay System combines two existing technologies: (1) a radiological imaging device for patient positioning and target monitoring and (2) a radiation therapy planning and delivery system. As requested by the FDA, this initial submission installment covers the treatment planning and delivery system software (TPDS) of the device. A subsequent submission installment will cover the hardware aspects of the ViewRay System and its integration with the TPDS.

The new, prescription - use ViewRay System device contains software, which is the subject of this submission, and emits radiation. The ViewRay System will consist of a treatment planning system, 0.35T magnetic resonance imager and radiation delivery system. The ViewRay System can use MRI - guided radiation to deliver stereotactic radiosurgery and precision radiotherapy for lesions, tumors, and conditions anywhere in the body when radiation treatment is indicated as prescribed by a physician. The system utilizes the radionuclide Cobalt 60 as the source of radiation, which is a well - established technology.

At the FDA pre - 510(k) meeting in October 2009 (I070011), the FDA requested that ViewRay submit a 510(k) Premarket Notification for the Treatment Planning and Delivery System software ahead of the MRI - guided radiation therapy system description it is to be used with.

The software system has two major roles: radiotherapy treatment planning, and radiotherapy treatment delivery.

The TPDS is designed to be capable of executing on both a remote treatment planning workstation and on the operator's console that also controls the treatment delivery system. On remote workstations, treatment plans may be created to pre - plan treatment using volumetric computed tomography (CT), magnetic resonance imaging (MRI), or positron emission tomography (PET) images.

When the TPDS software executes on the operator's console, it allows for on - table planning. Treatment plans can be modified or updated before treatment delivery commences or in - between delivered fractions with proper authorization.

The software system is arranged such that the radiotherapy treatment planning system does not require any additional hardware except for the actual personal computer that the software is executing. Multiple planning stations can therefore be utilized separate from the treatment delivery system.

The radiotherapy delivery software is necessary to operate the hardware to deliver therapy. This software executes on the operator's console. The operator's console has both capabilities of treatment delivery and treatment planning.

A part of the treatment delivery system includes the MRI imaging system. This imaging system is developed by ViewRay as part of collaboration with Siemens Medical Systems. ViewRay's operator's console interfaces with the MRI imaging system; however, the details of this imaging system will be described in the hardware submission.

The software system also functions as an archival system to archive all patient data that is required for treatment planning and all data acquired during treatment delivery. This archival system is designed for use with the ViewRay System; however it will be able to support Digital Imaging and Communications in Medicine (DICOM) export of data to other third party archival systems.

	Yes	No	N/A
Is the device life-supporting or life sustaining?		X	
Is the device an implant (implanted longer than 30 days)?		X	
Does the device design use software?	X		
Is the device sterile?		X	
Is the device reusable (not reprocessed single use)?		X	
Are "cleaning" instructions included for the end user?			

IV. **Indications for Use:** The ViewRay Treatment Planning and Delivery System is indicated for the stereotactic radiosurgery and precision radiotherapy for lesions, tumors, and conditions anywhere in the body when radiation treatment is indicated.

V. **Predicate Device**: Varian Medical Systems Trilogy Mx System (K092871), Varian Eclipse Treatment Planning System (K091492). The Varian Medical Systems' Trilogy®/Eclipse® System is an integrated system that includes treatment-planning, imaging, and treatment delivery functions that are directly analogous to the functions provided by the ViewRay™ System. Both the Varian and ViewRay™ Systems can use images obtained from CT, PET or MRI for planning. Although the Varian system uses a different technology for obtaining images during delivery of treatment (CT vs. MRI) and as a source of radiation (Linac vs. Cobalt-60), both systems have the same intended use and indications for use, and are used by the same user population. The differences are as follows:

1. Cobalt-60 Photon beam dose computation and display.
2. Monte Carlo dose computation algorithm.
3. Deformable registration for tracking and adapting to interfractional changes during treatment.
4. Delivery gating capability*
5. On board imaging for volumetric and planar acquisition*

VI. **Labeling:** The firm provided the device labels, the device user manual, and IFU. There are no problems with the Labeling.

VII. Sterilization/Shelf Life/Reuse: NA.

VIII. Biocompatibility: NA

IX. Software: Mr. Steven Pelham, software specialist, has reviewed this part for the major level for concern of the software. Foe a total 11 major concerning of the software; he has no any questions. (See attached review, OK).

Version: 2.19.1.3

Level of Concern: Major

	Yes	No
Software description:	X	
Device Hazard Analysis:	X	
Software Requirements Specifications:	X	
Architecture Design Chart:	X	
Design Specifications:	X	
Traceability Analysis/Matrix:	X	
Development:	X	
Verification & Validation Testing:	X	
Revision level history:	X	
Unresolved anomalies:	X	

V. **Electromagnetic Compatibility and Electrical, Mechanical and Thermal Safety:** NA

XI. Substantial Equivalence Discussion

	Yes	No	
1. Same Indication Statement?	X		If **YES** = Go To 3
2. Do Differences Alter The Effect Or Raise New Issues of Safety Or Effectiveness?			If **YES** = Stop **NSE**
3. Same Technological Characteristics?		X	If **YES** = Go To 5
4. Could The New Characteristics Affect Safety Or Effectiveness?		X	If **YES** = Go To 6
5. Descriptive Characteristics Precise Enough?			If **NO** = Go To 8 If **YES** = Stop **SE**
6. New Types Of Safety Or Effectiveness Questions?		X	If **YES** = Stop **NSE**
7. Accepted Scientific Methods Exist?	X		If **NO** = Stop **NSE**
8. Performance Data Available?	X		If **NO** = Request Data
9. Data Demonstrate Equivalence?	X		Final Decision: SE

ViewRay has provided evidence that their device is SE to previously marketed device. All issues raised during review of this submission have been addressed.

ADDITIONAL SUPPORTING INFORMATION: None.

RECOMMENDATION:

I believe that this device is SE to: 90- MUJ
Classification should be based on: 892.5750 Radiation Therapy Treatment Planning System
Class: II (final)

John Chen, Ph. D. 01/05/11
Reviewer Date

Branch Chief 1/12/11
Date

K102915F
01/06/11

Memorandum of a

Software Review

K102915

January 4, 2011

From: Steven Pelham; OSEL-DESE (WO62-4229); (301) 796-2589
To: John Chen, OIVD-DRAD (HFZ-470), (240) 276-3643

Subject: Software review ViewRay Treatment Planning and Delivery System by ViewRay

---

Succinct Conclusion: Substantially Equivalent

The information contained within this submission is sufficient to meet the software concerns as described in the *Guidance for the Content of Premarket Submissions for Software Contained in Medical Devices*, dated May 11, 2005, and it is recommended that, from a software standpoint, this submission be considered **Substantially Equivalent**.

Summary

The ViewRay™ Treatment Planning and Delivery System is intended to be used for planning external beam irradiation with photon beams and delivering stereotactic radiosurgery and precision radiotherapy for lesions, tumors, and conditions anywhere in the body when radiation treatment is indicated, in conjunction with the ViewRay System, an MRI image - guided radiation therapy system.

The ViewRay Treatment Planning and Delivery System (TPDS) provides tools for planning and delivery of external gamma beam stereotactic radiosurgery and precision radiotherapy for lesions, tumors, and conditions anywhere in the body when radiation treatment is indicated. It is a computer - based device used by trained medical professionals. The Treatment Planning software is only designed to be used on the ViewRay radiation therapy system. The ViewRay TPDS is capable of assisting clinicians in creating treatment delivery and QA plans for the ViewRay System, and reviewing, prescribing, tracking, and correcting the course of patient treatment using tools for contouring, visualization, data storage, anatomical target monitoring and reoptimization.

The ViewRay System is an MR image - guided radiation therapy device. Similar to other existing image - guided radiation therapy (IGRT) systems in the marketplace, the ViewRay System combines two existing technologies: (1) a radiological imaging device for patient positioning and target monitoring and (2) a radiation therapy planning and delivery system. As requested by the FDA, this initial submission installment covers the treatment planning and delivery system software (TPDS) of the device. A subsequent submission installment will cover the hardware aspects of the ViewRay System and its integration with the TPDS.

The new, prescription - use ViewRay System device contains software, which is the subject of this submission, and emits radiation. The ViewRay System will consist of a treatment planning system, 0.35T magnetic resonance imager and radiation delivery system. The ViewRay System can use MRI - guided radiation to deliver stereotactic radiosurgery and precision radiotherapy for lesions, tumors, and conditions anywhere in the body when radiation treatment is indicated as prescribed by a physician. The system utilizes the radionuclide Cobalt 60 as the source of radiation, which is a well - established technology.

At the FDA pre - 510(k) meeting in October 2009 (I070011), the FDA requested that ViewRay submit a 510(k) Premarket Notification for the Treatment Planning and Delivery System software ahead of the MRI - guided radiation therapy system description it is to be used with.

The software system has two major roles: radiotherapy treatment planning, and radiotherapy treatment delivery.

The TPDS is designed to be capable of executing on both a remote treatment planning workstation and on the operator's console that also controls the treatment delivery system. On remote workstations, treatment plans may be created to pre - plan treatment using volumetric computed tomography (CT), magnetic resonance imaging (MRI), or positron emission tomography (PET) images.

When the TPDS software executes on the operator's console, it allows for on - table planning. Treatment plans can be modified or updated before treatment delivery commences or in - between delivered fractions with proper authorization.

The software system is arranged such that the radiotherapy treatment planning system does not require any additional hardware except for the actual personal computer that the software is executing. Multiple planning stations can therefore be utilized separate from the treatment delivery system.

The radiotherapy delivery software is necessary to operate the hardware to deliver therapy. This software executes on the operator's console. The operator's console has both capabilities of treatment delivery and treatment planning.

A part of the treatment delivery system includes the MRI imaging system. This imaging system is developed by ViewRay as part of a collaboration with Siemens Medical Systems. ViewRay's operator's console interfaces with the MRI imaging system; however, the details of this imaging system will be described in the hardware submission.

The software system also functions as an archival system to archive all patient data that is required for treatment planning and all data acquired during treatment delivery. This archival system is designed for use with the ViewRay System, however it will be able to support Digital Imaging and Communications in Medicine (DICOM) export of data to other third party archival systems.

This software review does not cover the correctness of the underlying algorithms or their appropriateness or applicability to the indicated use of this device, as such evaluation is beyond the scope of this review. This is a MAJOR Level of Concern device.

1. **Software Description [ok]**

   The firm provided an acceptable overview of the device features that are controlled by software, and a description of the intended operational environment. [Section 2B. Software Features & Operating Environment Overview]

2. **Device (including Software) Hazard Analysis [ok]**

   The firm provided an acceptable description of the hazards presented by this device, the causes and severity of the hazards, the method of control of the hazards and the testing done to verify the correct implementation of that method of control, and any residual hazards. [Section 2D, Device/Software Hazard Analysis; Section 4, Risk Control Discussion, Attachments 4D-2 & 4D-3, Risk Assessment & Control Plan - Software]

3. **Software Requirements Specifications (SRS) [ok]**

   The firm provided a copy of their software requirements specification document, which clearly documented their functional, performance, interface, design, and development requirements. [Attachment 2E-1, ViewRay System Software Requirements Specification]

4. **Architecture Design [ok]**

   The firm provided an acceptable description of the software system partitioned into its functional subsystems. [Attachment 2F-1, ViewRay System Software Architecture Design]

5. **Software Design Specification (SDS) [ok]**

   The firm provided an acceptable design specification document, which describes what the program does and how it does it. [Attachments 2G-1, User Interface Software Design Specification; 2G-2, Algorithm Engine SDS; 2G-3 Visualization SDS; 2G-4 Data Model SDS; 2G-5 Database SDS; 2G-6 Services SDS; 2G-7 Logging SDS]

6. **Traceability Analysis [ok]**

   The provided their traceability matrix, which provided the links between the hazards, requirements, validation and testing. [Attachment 6A-6, Traceability Matrix- Requirements V&V]

7. **Software Development Environment Description [ok]**

   The firm submitted a summary of their software development life cycle plan, describing the processes that have been put into place to manage the various software development life cycle activities, including a summary of the configuration management and maintenance activities. [Attachments 2H-1, Software Development; 2H-2, Software Development Plan, Rev F; 2H-3, Software Coding Standards; 2H-4, Software Configuration Management, Rev C; 2H-5, Design Issue Tracking; 2H-6, Corrective and Preventive Action; 2H-7, Design Change Control; 2H-8, Control of Nonconforming Product]

8. **Verification and Validation (including Testing) [ok]**

   The firm provided an acceptable description of their systematic process of life cycle activities, including analysis, evaluation, assurance, and testing of the software, and supporting documentation. [Attachment 6A-1, Algorithm Engine Unit Test Verification

Report; 6A-2 Treatment Planning Delivery Software Verification Validation Summary Report; 6A-3 Verification Test Protocol and Report (Requirements); 6A-4 Verification Test Protocol and Report (Hazard Controls); 6A-5 SW Revision 2.19.1.3 Regression Test Report;

9. Revision Level History [ok]

The firm provided the revision history log, documenting all major changes to the software during its development cycle. [Attachment 2I-1, Software Revision Level History & Release Version Number]

10. Unresolved Anomalies (bugs) [ok]

The firm provided a list of all unresolved software anomalies, indicating the problem, the impact on device performance, and plans and time frames for correcting each problem. [Attachment 2J-1, Software Unresolved Anomalies]

11. Release Version Number [ok]

The firm provided an acceptable description of the version number [2.19.1.3] and date [9.23.2010]. [Attachment 2I-1, Software Revision Level History & Release Version Number]

Recommendation — Substantially Equivalent

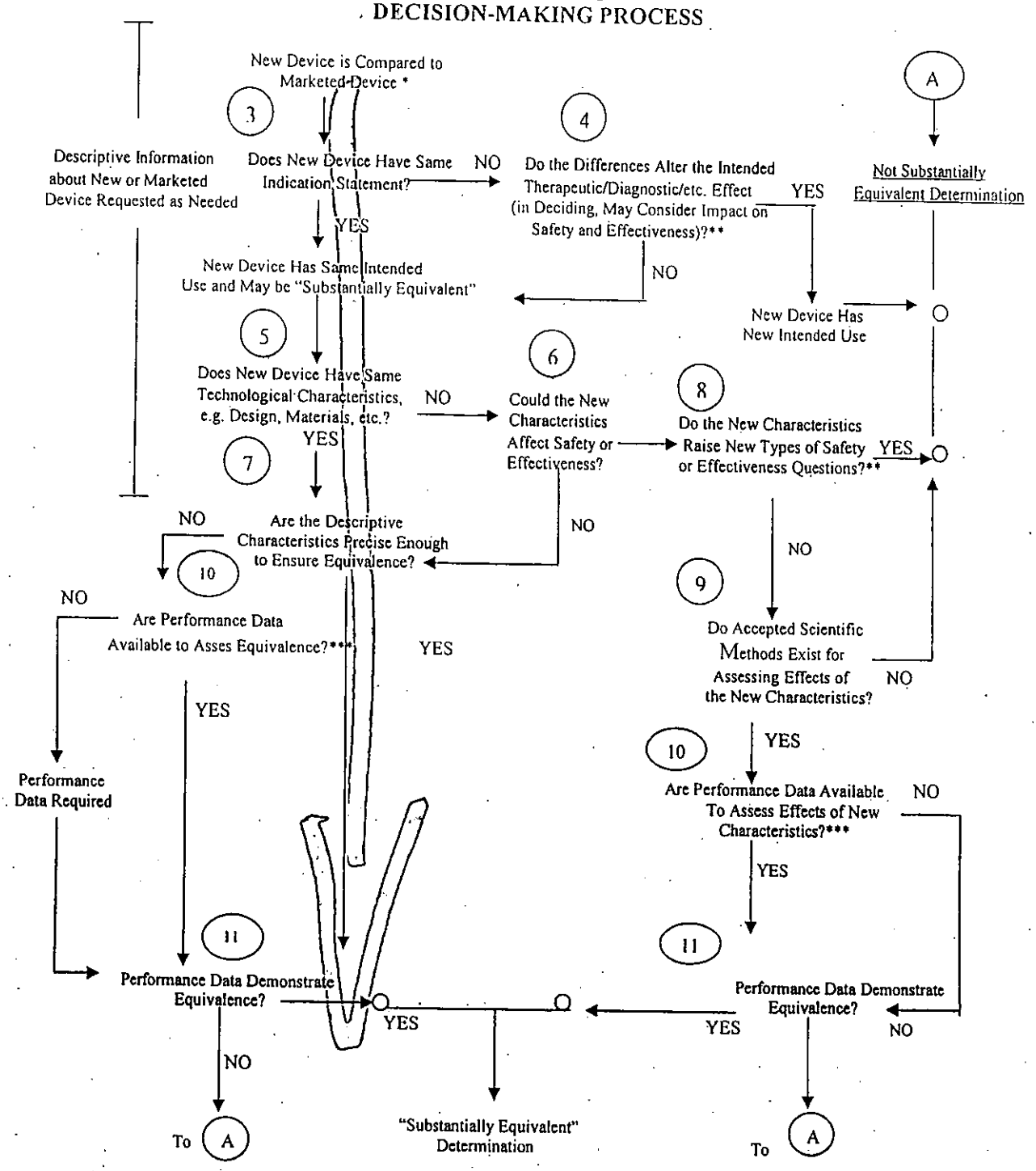
The firm has provided acceptable documentation demonstrating that they:

- have developed the software for this device under an appropriate software development program;
- have performed a hazard analysis from both the patient's and user's standpoint;
- addressed those hazards; and
- carried out an appropriate validation process.

These procedures provide the foundation for assuring, to the extent possible, that the software will operate in a manner described in the specifications, and in no other predictable way. It is recommended that, from a software standpoint, this submission should be considered **Substantially Equivalent**.

14

510(k) "SUBSTANTIAL EQUIVALENCE" DECISION-MAKING PROCESS



❖ 510(k) Submissions compare new devices to marketed devices. FDA requests additional information if the relationship between marketed and "predicate" (pre-Amendments or reclassified post-Amendments) devices is unclear.

❖❖ This decision is normally based on descriptive information alone, but limited testing information is sometimes required.

❖❖❖ Data maybe in the 510(k), other 510(k)s, the Center's classification files, or the literature.