Monte Carlo Techniques in Radiation Therapy

Edited by
Joao Seco
Frank Verhaegen
Monte Carlo Techniques in Radiation Therapy
Quality and Safety in Radiotherapy
Todd Pawlicki, Peter B. Dunscombe, Arno J. Mundt, and Pierre Scalliet, Editors
ISBN: 978-1-4398-0436-0

Adaptive Radiation Therapy
X. Allen Li, Editor

Quantitative MRI in Cancer
Thomas E. Yankeeelov, David R. Pickens, and Ronald R. Price, Editors

Informatics in Medical Imaging
George C. Kagadis and Steve G. Langer, Editors

Adaptive Motion Compensation in Radiotherapy
Martin J. Murphy, Editor
ISBN: 978-1-4398-2193-0

Image-Guided Radiation Therapy
Daniel J. Bourland, Editor
ISBN: 978-1-4398-0273-1

Targeted Molecular Imaging
Michael J. Welch and William C. Eckelman, Editors
ISBN: 978-1-4398-4195-0

Proton and Carbon Ion Therapy
C.-M. Charlie Ma and Tony Lomax, Editors
ISBN: 978-1-4398-1607-3

Comprehensive Brachytherapy: Physical and Clinical Aspects
Jack Venselaar, Dimos Baltas, Peter J. Hoskin, and Ali Soleimani-Meigooni, Editors
ISBN: 978-1-4398-4498-4

Physics of Mammographic Imaging
Mia K. Markey, Editor

Physics of Thermal Therapy: Fundamentals and Clinical Applications
Eduardo Moros, Editor
ISBN: 978-1-4398-4890-6

Emerging Imaging Technologies in Medicine
Mark A. Anastasio and Patrick La Riviere, Editors
ISBN: 978-1-4398-8041-8

Cancer Nanotechnology: Principles and Applications in Radiation Oncology
Sang Hyun Cho and Sunil Krishnan, Editors
ISBN: 978-1-4398-7875-0

Monte Carlo Techniques in Radiation Therapy
Joao Seco and Frank Verhaegen, Editors
ISBN: 978-1-4665-0792-0

Forthcoming titles in the series

Image Processing in Radiation Therapy
Kristy Kay Brock, Editor
ISBN: 978-1-4398-3017-8

Stereotactic Radiosurgery and Radiotherapy
Stanley H. Benedict, Brian D. Kavanagh, and David J. Schlesinger, Editors
ISBN: 978-1-4398-4197-6

Informatics in Radiation Oncology
Bruce H. Curran and George Starkschall, Editors

Tomosynthesis Imaging
Ingrid Reiser and Stephen Glick, Editors

Cone Beam Computed Tomography
Chris C. Shaw, Editor
ISBN: 978-1-4398-4626-1
To all who made this book possible
# Contents

Series Preface ............................................................................................................................................................................ ix
Preface ........................................................................................................................................................................................ xi
Editors ...................................................................................................................................................................................... xiii
Contributors ............................................................................................................................................................................ xv

## PART I  Monte Carlo Fundamentals

1 History of Monte Carlo ................................................................. 3
   Alex F. Bielajew
2 Basics of Monte Carlo Simulations ........................................ 17
   Matthias Fippel
3 Variance Reduction Techniques .................................................. 29
   Matthias Fippel

## PART II  Application of Monte Carlo Techniques in Radiation Therapy

4 Applications of Monte Carlo to Radiation Dosimetry .............. 43
   Hugo Bouchard and Jan Seuntjens
5 Monte Carlo Modeling of External Photon Beams in Radiotherapy ........ 63
   Frank Verhaegen
6 Monte Carlo Modeling of External Electron Beams in Radiotherapy ........................................................................... 87
   Frank Verhaegen
7 Dynamic Beam Delivery and 4D Monte Carlo ............................ 95
   Emily Heath and Joao Seco
8 Patient Dose Calculation ......................................................... 111
   Joao Seco and Maggy Fragoso
9 Monte Carlo Methods and Applications for Brachytherapy Dosimetry and Treatment Planning ........................................ 125
   Guillaume Landry, Mark J. Rivard, Jeffrey F. Williamson, and Frank Verhaegen
10 Monte Carlo as a QA Tool for Advanced Radiation Therapy .... 145
   JinSheng Li and C.-M. Charlie Ma
11 Electrons: Clinical Considerations and Applications ................ 155
   Joanna E. Cygler and George X. Ding
<table>
<thead>
<tr>
<th>Chapter</th>
<th>Title</th>
<th>Pages</th>
</tr>
</thead>
<tbody>
<tr>
<td>12</td>
<td>Photons: Clinical Considerations and Applications</td>
<td>167</td>
</tr>
<tr>
<td></td>
<td><em>Michael K. Fix</em></td>
<td></td>
</tr>
<tr>
<td>13</td>
<td>Monte Carlo Calculations for Proton and Ion Beam Dosimetry</td>
<td>185</td>
</tr>
<tr>
<td></td>
<td><em>Hugo Palmans</em></td>
<td></td>
</tr>
<tr>
<td>14</td>
<td>Protons: Clinical Considerations and Applications</td>
<td>201</td>
</tr>
<tr>
<td></td>
<td><em>Harald Paganetti</em></td>
<td></td>
</tr>
<tr>
<td>15</td>
<td>Application of Monte Carlo Methods to Radionuclide Therapy</td>
<td>223</td>
</tr>
<tr>
<td></td>
<td><em>Michael Ljungberg</em></td>
<td></td>
</tr>
<tr>
<td>16</td>
<td>Monte Carlo for Kilovoltage and Megavoltage Imaging</td>
<td>233</td>
</tr>
<tr>
<td></td>
<td><em>George X. Ding and Andrew Fielding</em></td>
<td></td>
</tr>
<tr>
<td>17</td>
<td>Monte Carlo Calculations for PET-Based Treatment Verification of Ion Beam Therapy</td>
<td>251</td>
</tr>
<tr>
<td></td>
<td><em>Katia Parodi</em></td>
<td></td>
</tr>
<tr>
<td>18</td>
<td>Monte Carlo Studies of Prompt Gamma Emission and of Proton Radiography/Proton-CT</td>
<td>263</td>
</tr>
<tr>
<td></td>
<td><em>Jerimy C. Polf, Nicholas Depauw, and Joao Seco</em></td>
<td></td>
</tr>
<tr>
<td>19</td>
<td>Monte Carlo for Treatment Device Design</td>
<td>273</td>
</tr>
<tr>
<td></td>
<td><em>Bruce A. Faddegon</em></td>
<td></td>
</tr>
<tr>
<td>20</td>
<td>GPU-Based Fast Monte Carlo Simulation for Radiotherapy Dose Calculation</td>
<td>283</td>
</tr>
<tr>
<td></td>
<td><em>Xun Jia, Sami Hissoiny, and Steve B. Jiang</em></td>
<td></td>
</tr>
<tr>
<td>21</td>
<td>Monte Carlo for Shielding of Radiotherapy Facilities</td>
<td>295</td>
</tr>
<tr>
<td></td>
<td><em>Peter J. Biggs and Stephen F. Kry</em></td>
<td></td>
</tr>
</tbody>
</table>
Advances in the science and technology of medical imaging and radiation therapy are more profound and rapid than ever before, since their inception over a century ago. Further, the disciplines are increasingly cross-linked as imaging methods become more widely used to plan, guide, monitor, and assess treatments in radiation therapy. Today the technologies of medical imaging and radiation therapy are so complex and so computer driven that it is difficult for the persons (physicians and technologists) responsible for their clinical use to know exactly what is happening at the point of care, when a patient is being examined or treated. The persons best equipped to understand the technologies and their applications are medical physicists, and these individuals are assuming greater responsibilities in the clinical arena to ensure that what is intended for the patient is actually delivered in a safe and effective manner.

The growing responsibilities of medical physicists in the clinical arenas of medical imaging and radiation therapy are not without their challenges, however. Most medical physicists are knowledgeable in either radiation therapy or medical imaging, and expert in one or a small number of areas within their discipline. They sustain their expertise in these areas by reading scientific articles and attending scientific talks at meetings. In contrast, their responsibilities increasingly extend beyond their specific areas of expertise. To meet these responsibilities, medical physicists periodically must refresh their knowledge of advances in medical imaging or radiation therapy, and they must be prepared to function at the intersection of these two fields. How to accomplish these objectives is a challenge.

At the 2007 annual meeting of the American Association of Physicists in Medicine in Minneapolis, this challenge was the topic of conversation during a lunch hosted by Taylor & Francis and involving a group of senior medical physicists (Arthur L. Boyer, Joseph O. Deasy, C.-M. Charlie Ma, Todd A. Pawlicki, Ervin B. Podgorsak, Elke Reitzel, Anthony B. Wolbarst, and Ellen D. Yorke). The conclusion of this discussion was that a book series should be launched under the Taylor & Francis banner, with each volume in the series addressing a rapidly advancing area of medical imaging or radiation therapy of importance to medical physicists. The aim would be for each volume to provide medical physicists with the information needed to understand technologies driving a rapid advance and their applications to safe and effective delivery of patient care.

Each volume in the series is edited by one or more individuals with recognized expertise in the technological area encompassed by the book. The editors are responsible for selecting the authors of individual chapters and ensuring that the chapters are comprehensive and intelligible to someone without such expertise. The enthusiasm of volume editors and chapter authors has been gratifying and reinforces the conclusion of the Minneapolis luncheon that this series of books addresses a major need of medical physicists.

Imaging in Medical Diagnosis and Therapy would not have been possible without the encouragement and support of the series manager, Luna Han of Taylor & Francis. The editors and authors, and most of all I, are indebted to her steady guidance of the entire project.

William Hendee
Series Editor
Rochester, Minnesota
Monte Carlo simulation techniques made a slow entry in the field of radiotherapy in the late 1970s. Since then they have gained enormous popularity, judging by the number of papers published and PhDs obtained on the topic. Calculation power has always been an issue, so initially only simple problems could be addressed. They led to insights, though, that could not have been obtained by any other method. Recently, fast forwarding some 30 years, Monte Carlo–based treatment planning tools have now begun to be available from some commercial treatment planning vendors, and it can be anticipated that a complete transition to Monte Carlo–based dose calculation methods may take place over the next decade. The progress of image-guided radiotherapy further advances the need for Monte Carlo simulations, in order to better understand and compute radiation dose from imaging devices and make full use of the four-dimensional information now available. Exciting new developments in in-beam imaging in light ion beams are now also being vigorously investigated. Many new discoveries await the use of the Monte Carlo technique in radiotherapy in the coming decades.

The book addresses the application of the Monte Carlo particle transport simulation technique in radiation therapy, mostly focusing on external beam radiotherapy and brachytherapy. It includes a presentation of the mathematical and technical aspects of the technique in particle transport simulations. It gives practical guidance relevant for clinical use in radiation therapy, discussing modeling of medical linacs and other irradiation devices, issues specific to electron, photon, proton/particle beams, and brachytherapy, utilization in the optimization of treatment planning, radiation dosimetry, and quality assurance (QA).

We have assembled this book—a first of its kind—to be useful to clinical physicists, graduate students, and researchers who want to learn about the Monte Carlo method; we hope you will benefit from the collective knowledge presented herein. In addition, the editors wish to sincerely thank all the outstanding contributing authors, without whom this work would not have been possible. They would also like to thank the series editor, Dr. Bill Hendee, for the opportunity. We sincerely thank the editorial and production staff at Taylor & Francis for the smooth collaboration and for the pleasant interactions.
Editors

Joao Seco is assistant professor of radiation oncology at Harvard Medical School and Massachusetts General Hospital in Boston, Massachusetts. He earned his PhD from the Institute of Cancer Research, University of London, UK in 2002. He held research positions at the Royal Marsden Hospital, UK and the Harvard Medical School, Boston, Massachusetts for several years. His group has published more than 50 research papers and he was the recipient of the Harvard Club of Australia Foundation Award in 2012 on the development of Monte Carlo and optimization techniques for use in radiation therapy of lung cancer. His interests range from proton imaging and therapy to photon beam modeling Monte Carlo, electronic portal imaging, and 4D Monte Carlo proton and photon dosimetry. He started working on Monte Carlo simulations while a master’s student at the Laboratory of Instrumentation and Experimental Particle Physics (LIP), a Portuguese research institute part of the CERN (the European Organization for Nuclear Research) worldwide network for particle physics research.

Frank Verhaegen is head of clinical physics research at the Maastro Clinic in Maastricht, the Netherlands. He holds a professorship at the University of Maastricht. Formerly, he was an associate professor at McGill University in Montréal, Canada. He earned his PhD from the University of Ghent in Belgium in 1996. He held research positions at the Royal Marsden Hospital and the National Physical Laboratory, UK for several years. Dr. Verhaegen is a Fellow of the Institute of Physics and Engineering in Medicine and the Institute of Physics. His group has published more than 120 research papers and was the recipient of the Sylvia Fedoruk Prize for best Canadian Medical Physics paper in 2007. His interests range broadly in imaging and dosimetry for photon and electron therapy, brachytherapy, particle therapy, and small animal radiotherapy. Dr. Verhaegen has been passionate about Monte Carlo simulations since the days of his master’s thesis in the late 1980s.
Contributors

Alex F. Bielajew  
Department of Nuclear Engineering and Radiological Sciences  
University of Michigan  
Ann Arbor, Michigan

Peter J. Biggs  
Department of Radiation Oncology  
Massachusetts General Hospital  
Boston, Massachusetts

Hugo Bouchard  
Centre hospitalier de l’Université de Montréal  
Montréal, Québec, Canada

Joanna E. Cygler  
Department of Medical Physics  
The Ottawa Hospital Cancer Centre  
Ottawa, Ontario, Canada

Nicholas Depauw  
Department of Radiation Oncology  
Massachusetts General Hospital  
Boston, Massachusetts

George X. Ding  
Department of Radiation Oncology  
Vanderbilt University School of Medicine  
Nashville, Tennessee

Bruce A. Faddegon  
Radiation Oncology Department  
University of California, San Francisco  
San Francisco, California

Andrew Fielding  
School of Physical and Chemical Sciences  
Queensland University of Technology  
Queensland, Australia

Matthias Fippel  
Brainlab AG  
Bavaria, Germany

Michael K. Fix  
Division of Medical Radiation Physics  
Inselspital  
University of Bern  
Bern, Switzerland

Maggy Fragoso  
Department of Communication and Strategy  
Alfa-Comunicações  
Praia, Santiago Island, Cape Verde

Emily Heath  
Department of Physics  
Ryerson University  
Toronto, Ontario, Canada

Sami Hissoiny  
Département de Génie Informatique et Génie Logiciel  
École Polytechnique de Montréal  
Montréal, Québec, Canada

Xun Jia  
Department of Radiation Medicine and Applied Sciences  
University of California, San Diego  
La Jolla, California

Steve B. Jiang  
Department of Radiation Medicine and Applied Sciences  
University of California, San Diego  
La Jolla, California

Stephen F. Kry  
Department of Radiation Physics  
MD Anderson Cancer Center  
University of Texas  
Houston, Texas

Guillaume Landry  
Maastro Clinic  
Maastricht, the Netherlands

JinSheng Li  
Fox Chase Cancer Center  
Philadelphia, Pennsylvania

Michael Ljungberg  
Department of Medical Radiation Physics  
Lund University  
Lund, Sweden

C.-M. Charlie Ma  
Fox Chase Cancer Center  
Philadelphia, Pennsylvania

Harald Paganetti  
Department of Radiation Oncology  
Massachusetts General Hospital and Harvard Medical School  
Boston, Massachusetts

Hugo Palmans  
Division of Acoustics and Ionising Radiation  
National Physical Laboratory  
Teddington, United Kingdom

Katia Parodi  
Heidelberg Ion Beam Therapy Center  
Heidelberg, Germany

Jerimy C. Polf  
Department of Physics  
Oklahoma State University  
Stillwater, Oklahoma
Mark J. Rivard
Department of Radiation Oncology
Tufts University School of Medicine
Boston, Massachusetts

Joao Seco
Department of Radiation Oncology
Harvard Medical School
Boston, Massachusetts

Jan Seuntjens
Medical Physics Unit
McGill University
Montréal, Québec, Canada

Frank Verhaegen
Maastro Clinic
Maastricht, the Netherlands

Jeffrey F. Williamson
VCU Massey Cancer Center
Richmond, Virginia
1

History of Monte Carlo

1.1 Motivating Monte Carlo

Generally speaking, the Monte Carlo method provides a numerical solution to a problem that can be described as a temporal evolution (“translation/reflection/mutation”) of objects (“quantum particles” [photons, electrons, neutrons, protons, charged nuclei, atoms, and molecules], in the case of medical physics) interacting with other objects based upon object–object interaction relationships (“cross sections”). Mimicking nature, the rules of interaction are processed randomly and repeatedly, until numerical results converge usefully to estimated means, moments, and their variances. Monte Carlo represents an attempt to model nature through direct simulation of the essential dynamics of the system in question. In this sense, the Monte Carlo method is, in principle, simple in its approach—a solution to a macroscopic system through simulation of its microscopic interactions. Therein is the advantage of this method. All interactions are microscopic in nature. The geometry of the environment, so critical in the development of macroscopic solutions, plays little role except to define the local environment of objects interacting at a given place at a given time.

The scientific method is dependent on observation (measurement) and hypothesis (theory) to explain nature. The conduit between these two is facilitated by a myriad of mathematical, computational, and simulation techniques. The Monte Carlo method exploits all of them. Monte Carlo is often seen as a “competitor” to other methods of macroscopic calculation, which we will call deterministic and/or analytic methods. Although the proponents of either method sometimes approach a level of fanaticism in their debates, a practitioner of science should first ask, “What do I want to accomplish?” followed by “What is the most efficient way to do it?” and then, “What serves science the best?” Sometimes the correct answer will be “Deterministic,” and other times it will be “Monte Carlo.” The most successful scientist will avail himself or herself of more than one method of approach.

There are, however, two inescapable realities. The first is that macroscopic theory, particularly transport theory, provides deep insight and allows one to develop sophisticated intuition as to how macroscopic particle fields can be expected to behave. Monte Carlo cannot compete very well with this. In discovering the properties of macroscopic field behavior, Monte Carlo practitioners operate very much like experimentalists. Without theory to provide guidance, discovery is made via trial and error, guided perhaps, by some brilliant intuition.

However complexity is measured, when it comes to developing an understanding of a physical problem, Monte Carlo techniques become, at some point, the most advantageous. A proof is given, in the appendix of this chapter, that the Monte Carlo method is more advantageous in the evolution of five and higher dimensional systems. The dimensionality is just one measure of a problem’s “complexity.” The problems in RTP (radiotherapy target practice) and dosimetry are typically of dimension 6.ε, or 7.ε. That is, particles move in Cartesian space, with position $\bar{x}$, that varies continuously, except at particle inception or expiration. They move with momentum, $\bar{p}$, that varies both discretely.
and continuously. The dimension of time is usually ignored for static problems, though it cannot be for nonlinear problems, where a particle’s evolution can be affected by the presence of other particles in the simulation. (The “space-charge” effect is a good example of this.) Finally, the \( \epsilon \) is a discrete dimension that can encompass different particle species, as well as intrinsic spin.

This trade-off, between complexity and time to solution is expressed in Figure 1.1.

Although the name “Monte Carlo method” was coined in 1947, at the start of the computer age, stochastic sampling methods were known long before the advent of computers. The first reference known to this author, is that of Comte de Buffon (1777) who proposed a Monte Carlo-like method to determine the outcome of an “experiment” consisting of repeatedly tossing a needle onto a ruled sheet of paper, to determine the probability of the needle crossing one of the lines. This reference goes back to 1777, well before the contemplation of automatic calculating machines. Buffon further calculated that a needle of length \( L \) tossed randomly on a plane ruled with parallel lines, distance \( d \) apart, where \( d > L \), would have a probability of crossing one of the rules lines of

\[ p = \frac{2L}{\pi d}. \]  

(1.1)

Much later, Laplace (1886) suggested that this procedure could be employed to determine the value of \( \pi \), albeit slowly. Several other historical uses of Monte Carlo predating computers are cited by Kalos and Whitlock (2008).

The idea of using stochastic sampling methods first occurred to Ulam,† who, while convalescing from an illness, played sol-itaire repeatedly, and then wondered if he could calculate the probability of success by combinatorial analysis. It occurred to him, that it would be possible to do so by playing a large number of games, tallying the number of successful plays (Metropolis, 1987; Eckhart, 1987), and then estimating the probability of success. Ulam communicated this idea to von Neumann who, along with Ulam and Metropolis, were working on theoretical calculations related to the development of thermonuclear weapons. Precise calculations of neutron transport are essential in the design of thermonuclear weapons. The atomic bomb was designed by experiments, mostly, with modest theoretical support. The trigger for a thermonuclear weapon is an atomic bomb, and the instrumentation is destroyed before useful signals can be extracted.† Von Neumann was especially intrigued with the idea. The modern Monte Carlo age was ushered in later, when the first documented suggestion of using stochastic sampling methods applied to radiation transport calculations appeared in correspondence between von Neumann and Richtmyer (Metropolis, 1987; Eckhart, 1987), on March 11, 1947. (Richtmyer was the leader of the Theoretical Division at Los Alamos National Laboratories [LANL].) This letter suggested the use of LANL’s ENIAC computer to do the repetitive sampling. Shortly afterward, a more complete proposal was written (von Neumann and Richtmyer, 1947). Although this report was declassified as late as 1988, inklings of the method, referred to as a “mix of deterministic and random/stochastic processes,” started to appear in the literature, as published abstracts (Ulam and von Neumann, 1945, 1947). Then in 1949, Metropolis and Ulam published their seminal, founding paper, “The Monte Carlo Method” (Metropolis and Ulam, 1949), which was the first unclassified paper on the Monte Carlo methods, and the first to have the name, “Monte Carlo” associated with stochastic sampling.

Already by the 1949, symposia on the Monte Carlo methods were being organized, focusing primarily on mathematical techniques, nuclear physics, quantum mechanics, and general statistical analysis. A later conference, the Symposium on Monte Carlo Methods, held at the University of Florida in 1954 (Meyer, 1981) was especially important. There were 70 attendees, many of whom would be recognized as “founding fathers” by Monte Carlo practitioners in the radiological sciences. Twenty papers were presented, including two involving gamma rays, spanning 282 pages in the proceedings. This proceedings also includes a 95-page bibliography, a grand summary of the work-to-date, with many references having their abstracts and descriptions published in the proceedings.

The rest, to quote an overused expression, is history. It is interesting to note the wonderful irony: This mathematical method was created for destruction by means of the most terrible weapon in history, the thermonuclear bomb. Fortunately, this weapon has never been used in conflict. Rather, millions have benefited from the development of Monte Carlo methods

† The direct quote from Ulam’s autobiography (Ulam, 1991). (p. 196, 1991 edition): “The idea for what was later called the Monte Carlo method occurred to me when I was playing solitaire during my illness.”

† The book Dark Sun, by Richard Rhodes, is an excellent starting point for the history of that topic (Rhodes, 1988).
for medicine. That topic, at least a small subset of it, will occupy the rest of this chapter.

As of this writing, with the data from 2011 still incomplete, we have found that about 3,00,000 papers have been published on the Monte Carlo method. If we restrict this search to only those papers related to medicine, the number of publications is almost 30,000. The 10% contribution to the Monte Carlo method seems to be consistent over time, at least since 1970. That represents an enormous investment in human capital to develop this most useful tool. The temporal evolution of this human effort is shown in Figure 1.2. Before 2005, the growth in both areas appears exponential in nature. The total effort shows three distinct areas of slope, with sudden changes, currently unexplained, though it may be due to the sudden emergence of “vector” and “massively parallel” machines, and the increase in research associated with this fundamentally new computer architecture. The growth in the “Medicine” area has been constant.

Since 2005, both areas are statistically consistent with constant output, with the “Medicine” area leveling out, at greater than 2100 publications/year. It appears that this communication is being written at the pinnacle of this scientific endeavor!

1.2 Monte Carlo in Medical Physics

Every historical review has its biases, and the one employed here will restrict the discussion to the applications of radiotherapy and radiation dosimetry. Moreover, the focus will be on the development of electron Monte Carlo, for reasons explained in the following paragraph. There is an abundance of reviews on the use of Monte Carlo in medical physics. A few of the more recent ones that discuss radiotherapy physics and dosimetry are Andreo (1985, 1991), Mackie (1990), Rogers and Bielajew (1990), Ma and Jiang (1999), Verhaegen and Seuntjens (2003), Rogers (2006).

The rise of the use of linear electron accelerators (LINACs) for radiotherapy, also ushered in the need to develop Monte Carlo methods for the purpose of dose prediction and dosimetry. LINACs employed in radiotherapy provide an energetic and penetrating source of photons that enter deep into tissue, sparing the surface and attenuating considerably less rapidly than 60Co or 137Cs beams. Relativistic electrons have a range of about 1 cm for each 2 MeV of kinetic energy in water. At its maximum, starting with a pencil beam of electrons, the diameter of the electron energy deposition, pear-shaped “plume” is also about 1 cm per 2 MeV of initial kinetic energy. These dimensions are commensurate with the organs being treated, as well as the organs at risk. The treatment areas are heterogeneous, with differences in composition and density. Moreover, the instruments used to meter dose are even more diverse. It is true now, as it was then, that the Monte Carlo method provides the only prediction of radiometric quantities that satisfies the accuracy demand of radiotherapy.

Thus, the history of the utility of the Monte Carlo method in medical physics is inextricably tied to the development of Monte Carlo methods of electron transport in complex geometries and in the description of electromagnetic cascades. *

The first papers employing the Monte Carlo method using electron transport, were authored by Robert R. Wilson (1950, 1951, 1952, p. 261), who performed his calculations using a “spinning wheel of chance.” † Although apparently quite tedious, Wilson’s method was still an improvement over the analytic methods of the time—particularly in studying the average behavior and fluctuations about the average (Rossi, 1952). Hebbard and P. R. Wilson (1955) used computers to investigate electron straggling and energy loss in thick foils. The first use of an electronic digital computer in simulating high-energy cascades by Monte Carlo methods was reported by Butcher and Messel (1958, 1960), and independently by Varfolomeev and Svetlolobov (1959). These two groups collaborated in a much publicized work (Messel et al., 1962) that eventually led to an extensive set of tables describing the shower distribution functions (Messel and Crawford, 1970)—the so-called “shower book.”

* Certainly there are important applications in brachytherapy and imaging that ignore electron transport. However, we shall leave that description to other authors.
† R. R. Wilson is also acknowledged as the founder of proton radiotherapy (Wilson, 1946).
For various reasons, two completely different codes were written in the early-to-mid 1960s to simulate electromagnetic cascades. The first was written by Zerby and Moran (1962a,b, 1963) of the Oak Ridge National Laboratory, motivated by the construction of the Stanford Linear Accelerator Center. Many physics and engineering problems were anticipated as a result of high-energy electron beams showering in various devices and structures at that facility. This code had been used by Alsmiller and others (Alsmiller and Moran, 1966, 1967, 1968, 1969; Alsmiller and Barish, 1969, 1974; Alsmiller et al., 1974) for a number of studies since its development.

The second code was developed by Nagel (Nagel and Schlier, 1963; Nagel, 1964, 1965; Völkel, 1966) and several adaptations have been reported (Völkel, 1966; Nicoli, 1966; Burfeindt, 1967; Ford and Nelson, 1978). The original Nagel version, which Ford and Nelson called SHOWER1, was a FORTRAN code written for high-energy electrons (≤1000 MeV) incident upon lead in cylindrical geometry. Six significant electron and photon interactions (bremsstrahlung, electron–electron scattering, ionization-loss, pair-production, Compton scattering, and the photoelectric effect) plus multiple Coulomb scattering were accounted for. Except for annihilation, positrons and electrons were treated alike and were followed until they reached a cutoff energy of 1.5 MeV (total energy). Photons were followed down to 0.25 MeV. The cutoff energies were as low as or lower than those used by either Messel and Crawford or by Zerby and Moran. The availability of Nagel’s dissertation (1964) and a copy of his original shower program provided the incentive for Nicoli (Nicoli, 1966) to extend the dynamic energy range and flexibility of the code in order for it to be made available as a practical tool for the experimental physicist. It was this version of the code that eventually became the progenitor of the EGS (electron gamma shower) code systems (Ford and Nelson, 1978; Nelson et al., 1985; Bielajew and Rogers, 1987; Kawrakow and Rogers, 2000; Hirayama et al., 2005).

On a completely independent track, and apparently independent from the electromagnetic cascade community, was Berger’s e–γ code. It was eventually released to the public as ETRAN in 1968 (Berger and Seltzer, 1968), though it is clear that internal versions were being worked on at NBS (now NIST) (Seltzer, 1989) since the early 1960s, on the foundations laid by Berger’s landmark paper (Berger, 1963). The ETRAN code then found its way, being modified somewhat, into the Sandia codes, EZTRAN (Halbleib and Vandevender, 1971), EZTRAN2 (Halbleib and Vandevender, 1973), SANDYL (Colbert, 1973), TIGER (Halbleib and Vandevender, 1975), CYLTRAN (Halbleib and Vandevender, 1976), CYLTRANNM (Halbleib and Vandevender, 1977), CYLTRANP (unpublished), SPHERE (Halbleib, 1978), TIGERP (Halbleib and Morel, 1979), ACCEPT (Halbleib, 1980), ACCEPTTM (Halbleib et al., 1981), SPHEM (Miller et al., 1981), and finally the all-encompassing ITS (Halbleib and Mehllhorn, 1986; Halbleib et al., 1992) codes. The ITS electron transport code was incorporated into the MCNP (Monte Carlo N-particle) code at Version 4, in 1990 (Hendricks and Briesmeister, 1991). The MCNP code lays claim to being a direct descendant of the codes written by the originators of the Monte Carlo method, Fermi, von Neumann, Ulam, as well as Metropolis and Richtmyer (Briesmeister, 1986). Quoting directly:

Much of the early work is summarized in the first book to appear on Monte Carlo by Cashwell and Everett in 1957. Shortly thereafter the first Monte Carlo neutron transport code MCS was written, followed in 1967 by MCN. The photon codes MCC and MCPP were then added and in 1973 MCN and MCC were merged to form MCNG. The above work culminated in Version 1 of MCNP in 1977. The first two large user manuals were published by W. L. Thompson in 1979 and 1981. This manual draws heavily from its predecessors.

The first appearance of electron transport in MCNP occurred with Version 4, in 1990 (Hendricks and Briesmeister, 1991). After that time, MCNP became an important player in medical-related research, to be discussed later.

Berger’s contribution (1963) is considered to be the de facto founding paper (and Berger the founding father) of the field of Monte Carlo electron and photon transport. That article, 81 pages long, established a framework for the next generation of Monte Carlo computational physicists. It also summarized all the essential theoretical physics for Monte Carlo algorithm development. Moreover, Berger introduced a specialized method for electron transport. Electron transport and scattering, for medical physics, dosimetry, and many other applications, is subject to special treatment. Rather than modeling every discrete electron interaction (of the order of 109 for relativistic electrons), cumulative scattering theories, whereby 102–103 individual elastic and inelastic events are “condensed” into single “virtual” single-scattering events, thereby enabling a speedup by factors of hundreds, typically. Nelson, the originator of the EGS code system, is quoted as saying (W. R. Nelson, personal communication, 2011), “Had I known about Berger’s work, I may not have undertaken the work on EGS!”

As for general-purpose uses in medical-related fields, with multi-material, combinatorial geometries, the two historically dominant players in RTP/Dosimetry, are the EGS and MCNP codes, introduced above. In the last decade, GEANT (Brun et al., 1982; Allison et al., 2006) has also made significant contributions as well, presently equal in use to MCNP. A plot of the number of papers published, using these methods is charted.

---

‡ 1959, to be exact (Cashwell and Everett, 1959).

† The codes and manuals referred to in this quote appear to have vanished.

7 There are some very relevant, alternative approaches, that the reader should be aware of, namely FLUKA (Aarnio et al., 1984; Fasso et al., 2005; Battistoni et al., 2007) (that traces its roots to 1964 (Ranft, 1964)), and the Penelope code (Salvat and Parrellada, 1984; Baró et al., 1995). As of this writing, the number of papers produced using these codes in medical areas is about 240, about half that of MCNP.
in Figure 1.3. Once MCNP introduced electron transport, we see, from Figure 1.3, that usage of MCNP experienced exponential growth in its use in medical-related areas. That exponential growth ended in about 2000. Since then, both the EGS and MCNP code systems seem to be experiencing steady use, with GEANT still, arguably, on the increase. If one considers all of the nonmedical literature related to Monte Carlo, MCNP is undeniably the most cited Monte Carlo code, by about a factor of 7 over EGS.

It should be emphasized that these two code systems are very different in nature. EGS has specifically targeted the medical area, since 1984, though it has enjoyed some use in other areas of physics as well. Some features are genuinely unique, such as the tracking of separate electron spins, a feature introduced (Namito et al., 1993) in EGS5 (Hirayama et al., 2005), as well as Doppler broadening (Namito et al., 1994), both inclusions of great interest to those doing research in synchrotron radiation light sources. Overall, however, considering Monte Carlo uses over all areas, MNCP's state-of-the-art neutron transport, makes it the world leader in the nuclear and radiological sciences conglomerate. The EGS code systems are supported by practitioners representing a panoply of scientific disciplines: medical physics, radiological scientists, pure and applied physicists. MCNP serves these communities as well, but also enjoys the support of a vibrant nuclear engineering profession, where transport theory of neutrons is a rich and active research area.

1.3 EGSx Code Systems

The history of how EGS, a code developed primarily for high-energy physics shielding and detector simulations, came to be used in medical physics, has never appeared in text, in its entirety. Rogers' (2006) humility probably interfered with its exposition in his article. However, I was an observer at those early events, and think I may offer some insights. In 1978, SLAC published EGS3 (Ford and Nelson, 1978), and Rogers employed it in several important publications (Rogers, 1982, 1983a,b, 1984a,b). Of particular importance was the publication that offered a patch to the EGS3 algorithms to mimic a technique employed in ETRAN, making electron-dependent calculations reliable, by shortening the steps to virtual interactions. At the time, electron transport step-size artifacts were completely unexplained. Shortening the steps is understood to solve the problem, as explained most eloquently by Larsen (1992), but at the cost of added computational time. These "step-size artifacts," attracted the attention of Nelson, who invited Rogers to participate in authoring the next version of EGS, EGS4 (Nelson et al., 1985), along with Hirayama, a research scientist at the KEK, the High Energy Accelerator Research Organization, Tsukuba, Japan.

Following its release in December 1985, Rogers’ institution (a Radiation Standards’ Laboratory, and a sister laboratory to Berger’s NIST) became a nucleus of medical physics and dosimetry Monte Carlo dissemination. It took over support and distribution of the EGS4 code, and began offering training courses all over the world. Hirayama was engaged in similar efforts in the Asian regions.

Yet, the step-size artifacts in EGS4 remained unexplained. Nahum, who was interested in modeling ionization chamber response, visited Rogers’ laboratory in the spring of 1984, to collaborate on this topic. Nahum already had a scholarly past in electron Monte Carlo (Nahum, 1976), producing what would eventually be realized, through a Lewis (1950) moments analysis (Kawrakow and Bielajew, 1998), to be a far superior electron transport algorithm. While using EGS4 to predict ionization chamber response, EGS4 would predict responses that could be 60% too low. Quoting Nahum, “How could a calculation that one could sketch on the back of an envelope, and get correct to within 5%, be 60% wrong using Monte Carlo?” Step-size reduction solved the problem (Bielajew et al., 1985; Rogers et al., 1985), but the search for a resolution to step-size anomalies was commenced, resulting in the PRESTA algorithm (Bielajew and Rogers, 1987, 1989). The release of PRESTA was followed by the demonstration of various small, but important, shortcomings (Rogers, 1993; Foote and Smyth, 1995). There were improvements over the years (Bielajew and Kawrakow, 1997a,b,c, Kawrakow and Bielajew, 1998), eventually resulting in a revision of the EGS code, known as EGSnrc (Kawrakow and Rogers, 2000) and EGS5 (Hirayama et al., 2005). A PRESTA-like improvement of ETRAN (Seltzer, 1991; Kawrakow, 1996) was even developed.

1.4 Application: Ion Chamber Dosimetry

The founding paper for applying Monte Carlo methods to ionization chamber response is attributed to Bond et al. (1978), who, using an in-house Monte Carlo code, calculated ionization chamber response as a function of wall thickness, to $^{60}$Co
y-irradiation. While validating the EGS code for this application, it was found that the EGS code had fundamental algorithmic difficulties with this low-energy regime, as well as this application. The resolution of these difficulties, not patent in other general-purpose Monte Carlo codes, became of great interest to this author. While general improvements to electron transport ensued, the fundamental problem was quite subtle, and was eventually described elegantly by Foote and Smythe (1995). In a nutshell, the underlying algorithmic reason that was identified arose from electron tracks being stopped at material boundaries, where cross sections change. EGS used the partial electron path to model a deflection of the electron, from the accumulated scattering power. The result was a rare, but important effect, the spurious generation of fluence singularities.

The literature on ionization chamber dosimetry is extensive. A partial compilation of very early contributions is (Bielajew et al., 1985; Rogers et al., 1985; Andreo et al., 1986; Bielajew, 1990; Rogers and Bielajew, 1990; Ma and Nahum, 1991).

Presently, the calculation of ionization chamber corrections is a very refined enterprise, with results being calculated to better than 0.1%. The literature on this topic is summarized by Bouchard and Seuntjens in their chapter in this book, “Applications of Monte Carlo to radiation dosimetry.” That chapter also summarizes the contribution of Monte Carlo to dosimetry protocol and basic dosimetry data, some of the earliest applications of Monte Carlo to medicine.

1.5 Early Radiotherapy Applications

For brevity, only the earliest papers are cited in this section, and the reader is encouraged to employ the comprehensive reviews already cited earlier in this article. Some of the very early history of radiotherapy applications is rarely mentioned, and I have attempted to gather them here.

The Monte Carlo modeling of Cobalt-60 therapy units was first mentioned in the ICRU Report # 18 (1971). However, a more complete descriptive work followed somewhat later (Rogers et al., 1985; Han et al., 1987).

The modeling of LINAC therapy units was first accomplished by Petti et al. (1983a,b) and then, soon after by Mohan et al. (1985).

Photoneutron contamination from a therapy unit was first described by Ing et al. (1982), although the simulation geometry was simplified.

Mackie et al. pioneered the convolution method (Mackie et al., 1985) and then with other collaborators, generated the first database of “kernels” or “dose-spread arrays” for use in radiotherapy (Mackie et al., 1988). Independently, these efforts were being developed by Ahnesjö et al. (1987). These are still in use today.

Modeling of electron beams from medical LINACs was first accomplished by Teng et al. (1986), Hogstrom et al. (1986), and then Manfredotti et al. (1990).

An original plan to use Monte Carlo calculation for “target button to patient dose” was proposed by Mackie et al. (1990).

That effort became known as “the OMEGA (an acronym for Ottawa Madison Electron Gamma Algorithm) Project.” However, early on in that project, a “divide-and-conquer” approach was adopted, whereby the fixed machine outputs (“phase-space files”) were used as inputs to a patient-specific target (applicators and patient), to generate a full treatment plan. This bifurcation spawned two industries, treatment head modeling, of which the BEAM/EGSx code is the most refined (Rogers et al., 1995) and is the most cited paper in the “Web of knowledge” with “Monte Carlo” in the title, and “radiotherapy” as a topic. The second industry spawned by the OMEGA project was the development of fast patient-specific Monte Carlo-based dose calculation algorithms (Kawrakow et al., 1995, 1996; Sempau et al., 2000; Gardner et al., 2007). For more discussion on the current fast Monte Carlo methods in use, the reader is encouraged to see the excellent review by Spezi and Lewis (2008).

1.6 The Future of Monte Carlo

The first step in predicting the future is to look where one has been, extrapolate the process, thereby predicting the future. The second step in predicting the future is to realize that the first step involves some very specious, and problematic reasoning!

The progress of time, with the events that it contains, is intrinsically “catastrophic” in nature. A scientific discovery can be so earth-shattering, that new directions of research are spawned, while others dissolve into irrelevance. Yet, we persist in the practice of prediction. Therefore, allow me to be very modest in this effort.

Amdahl’s Law (1967): Multiprocessor intercommunication bottlenecks will continue to be a limiting factor, for massively parallel machines. However, gains in traditional, monolayer, single chip speeds do not appear to be slowing, following Moore’s Law.∗

Harder to predict is algorithm development, specific to Monte Carlo applications in RTP. There is historical precedent for this in the citation data. In 1991, there was a 2.8 factor increase in productivity in only one year, followed by another in 1998, by a factor of 1.6. These increases are large, and unexplained. Yet, they illustrate the chaotic nature of the field.

There is a strong suggestion that research in Monte Carlo is saturating. Although new areas of research may rise and grow, both the “medical” and “all” data have been flat since about 2005. The EGS and MCNP output has been flat since 2000. Perhaps we are at the pinnacle?

We can safely predict, while the approach to the pinnacle may have been somewhat chaotic, the decline will be gradual. Monte Carlo codes have gotten easier to use, packaged with more

∗ In 1965, Gordon E. Moore, co-founder of Intel Corporation, predicted that areal transistor density would double every 18 months. One can infer a commensurate increase in computer speed. Moore changed his prediction to “doubling every two years” in 1975. Current chips designs follow Moore’s Law until 2015 at least.
user-friendly interfaces, and developing into “shrink-wrapped,” “turn-key” software systems. This is how it should be. The fact that a Monte Carlo method is the “engine” beneath a computational algorithm, should be transparent, but not invisible, to the researcher using the tool.

Paraphrasing the comment made by Martin Berger (Seligman et al., 1991), the founder of our field, during his speech at his retirement Festschrift symposium (1991),

I am not used to so much public attention. Tonight is quite unusual for me. I hope that, after tonight, I can disappear into the anonymity that I so assuredly crave.

And so it may be, for Monte Carlo research—at least, until the next great thing comes along.

A very astute student once said to me, “We no longer have Departments of Railroad Engineering,” in an effort to explain the (then) decline of Nuclear Engineering Departments, before the “nuclear renaissance.” It may be that the development of the Monte Carlo method is bound to decline as it matures, but it will remain an essential component of our scientific infrastructure, forever.

Appendix: Monte Carlo and Numerical Quadrature

In this appendix, we present a mathematical proof that the Monte Carlo method is the most efficient way of estimating tallies in three spatial dimensions when compared to first-order deterministic (analytic, phase-space evolution) methods. Notwithstanding the opinion that the Monte Carlo method is thought of as providing the most accurate calculation, the argument may be made in such a way, that it is independent of the physics content of the underlying algorithm or the quality of the incident radiation field.

A.1 Dimensionality of Deterministic Methods

For the purposes of estimating tallies from initiating electrons, photons, or neutrons, the transport process that describes the trajectories of particles is adequately described by the linear Boltzmann transport equation (Duderstadt and Martin, 1979):

\[
\frac{\partial}{\partial s} + \left( \frac{p}{|p|} \right) \cdot \nabla + \mu(x,p) \psi(x,p,s) = \int dx' \int dp' \mu(x,p,p') \psi(x',p',s),
\]

where \( x \) is the position, \( p \) is the momentum of the particle, \( (p/|p|) \cdot \nabla \) is a directional derivative (in three dimensions \( \Omega \cdot \nabla \), for example) and \( s \) is a measure of the particle pathlength. We use the notation that \( x \) and \( p \) are multi-dimensional variables of dimensionality \( N_x \) and \( N_p \). Conventional applications span the range \( 1 \leq N_p \leq 3 \). The macroscopic differential scattering cross section (probability per unit length) \( \mu(x,p,p') \) describes scattering from momentum \( p' \) to \( p \) at location \( x \), and the total macroscopic cross section is defined by

\[
\mu(x,p) = \int dp' \mu(x,p,p').
\]

\[\psi(x,p,s) \, dx \, dp \] is the probability of there being a particle in \( dx \) about \( x \), in \( dp \) about \( p \) and at pathlength \( s \). The boundary condition to be applied is

\[
\psi(x,p,0) = \delta(x)\delta(p_0 - p)\delta(s),
\]

where \( p_0 \) represents the starting momentum of a particle at \( s = 0 \). The essential feature of Equation A.1, insofar as this proof is concerned, is that the solution involves the computation of a \( (N_x + N_p) \)-dimensional integral.

A general solution may be stated formally:

\[
\psi(x,p,s) = \int dx' \int dp' G(x,p,x',p',s)Q(x',p'),
\]

where \( G(x,p,x',p',s) \) is the Green’s function and \( Q(x',p') \) is a source. The Green’s function encompasses the operations of transport (drift between points of scatter, \( x' \to x \)), scattering (i.e., change in momentum) and energy loss, \( p' \to p \). The interpretation of \( G(x,p,x',p',s) \) is that it is an operator that moves particles from one point in \( (N_x + N_p) \)-dimensional phase space, \( (x',p') \), to another, \( (x,p) \) and can be computed from the kinematical and scattering laws of physics.

Two forms of Equation A.4 have been employed extensively for general calculation purposes. Convolution methods integrate Equation A.4 with respect to pathlength \( s \) and further assume (at least for the calculation of the Green’s function) that the medium is effectively infinite. Thus,

\[
\psi(x,p) = \int dx' \int dp' G\left( \left| x - x' \right|, \left| \frac{p}{|p|} - \frac{p'}{|p'|} \right| \right) Q(x',p'),
\]

where the Green’s function is a function of the distance between the source point \( x' \) and \( x \), the angle between the vector defined by the source \( p' \) and \( p \) and the magnitude of the momentum of the course, \( |p'| \), or equivalently, the energy.

To estimate a tally using Equation A.5 we integrate \( \psi(x,p) \) over \( p \), with a response function, \( R(x,p) \) (Shultis and Faw, 1996):

\[
T(x) = \int dx' \int dp' R\left( \left| x - x' \right|, \left| p' \right| \right) Q(x',p'),
\]
where the “kernel,” \( F(|x - x'|, p') \), is defined by

\[
F(|x - x'|, p') = \int dp' \mathcal{R}(x, p') G\left(|x - x'|, \left| \frac{p}{p'} \right| \right),
\]

(A.7)

\( F(|x - x'|, p') \) has the interpretation of a functional relationship that connects particle fluence at phase-space location \( x', p' \) to a tally calculated at \( x \). This method has a known difficulty—its treatment of heterogeneities and interfaces. Heterogeneities and interfaces can be treated approximately by scaling \( |x - x'| \) by the collision density. This is an exact for the part of the kernel that describes the first scatter contribution but approximate for higher-order scatter contributions. It can also be approximate, to varying degrees, if the scatter produces other particles with different scaling laws, such as the electron set in motion by a first Compton collision.

For calculation methods that are concerned with primary charged particles, the heterogeneity problem is more severe. The true solution in this case is reached when the path length steps, \( s \) in Equation A.4 are made small (Larsen, 1992) and so, an iterative scheme is set up:

\[
\begin{align*}
\psi_1(x, p) &= \int dx' \int dp' G(x, p, x', p', \Delta s) Q(x', p') \\
\psi_2(x, p) &= \int dx' \int dp' G(x, p, x', p', \Delta s) \psi_1(x', p') \\
\psi_3(x, p) &= \int dx' \int dp' G(x, p, x', p', \Delta s) \psi_2(x', p') \\
&\quad \quad \quad \quad \vdots \\
\psi_N(x, p) &= \int dx' \int dp' G(x, p, x', p', \Delta s) \psi_{N-1}(x', p')
\end{align*}
\]  

(A.8)

which terminates when the largest energy in \( \psi_N(x, p) \) has fallen below an energy threshold or there is no \( x \) remaining within the target. The picture is of the phase space represented by \( \psi(x, p) \) “evolving” as \( s \) accumulates. This technique has come to be known as the “phase-space evolution” model. Heterogeneities are accounted for by forcing \( \Delta s \) to be “small” or of the order of the dimensions of the heterogeneities and using a \( G() \) that pertains to the atomic composition of the local environment. The calculation is performed in a manner similar to the one described for convolution. That is,

\[
T(x) = \sum_{j=1}^{N} \int dx' \int dp' F(x, x', p, p', \Delta s) \psi_j(x', p'),
\]

(A.9)

where the “kernel,” \( F(x, x', p, p', \Delta s) \), is defined by

\[
F(x, x', p, p', \Delta s) = \int dp \mathcal{R}(x, p) G(x, p, x', p', \Delta s).
\]

(A.10)

In the following analysis, we will not consider further any systematic errors associated with the treatment of heterogeneities in the case of the convolution method, nor with the “stepping errors” associated with incrementing \( s \) using \( \Delta s \) in the phase space evolution model. Furthermore, we assume that the Green’s functions or response kernels can be computed “exactly”—that there is no systematic error associated with them. The important result of this discussion is to demonstrate that the dimensionality of the analytic approach is \( N_i + N_p \).

### A.2 Convergence of Deterministic Solutions

The discussion of the previous section indicates that deterministic solutions are tantamount to solving a D-dimensional integral of the form:

\[
I = \int du H(u).
\]

(A.11)

In \( D \) dimensions, the calculation is no more difficult than in two dimensions, only the notation is more cumbersome. One notes that the integral takes the form:

\[
I = \int du_1 \int du_2 \cdots \int du_D H(u_1, u_2, \ldots, u_D)
\]

\[
= \sum_{r_1=1}^{N_1} \int du_1 \sum_{r_2=1}^{N_2} \int du_2 \cdots \int du_D H(u_1, u_2, \ldots, u_D)
\]

\[
\times \sum_{r_D=1}^{N_D} H(u_1, u_2, \ldots, u_D)
\]

(A.12)

The Taylor expansion takes the form

\[
H(u_1, u_2, \ldots, u_D) = H(u_1, u_2, \ldots, u_D) + \sum_{j=1}^{D} (u_j - u_{j_0}) \]

\[
\times \partial H(u_1, u_2, \ldots, u_D)/\partial u_j + \sum_{j=1}^{D} \frac{(u_j - u_{j_0})^2}{2}
\]

\[
\times \partial^2 H(u_1, u_2, \ldots, u_D)/\partial u_j \partial u_j + \sum_{j=1}^{D} \sum_{k=j+1}^{D} (u_j - u_{j_0})(u_k - u_{k_0})
\]

\[
\times \partial^2 H(u_1, u_2, \ldots, u_D)/\partial u_j \partial u_k
\]

(A.13)
The linear terms of the form \((u_i - u_j)\) and the bilinear terms of the form \((u_i - u_j)(u_k - u_l)\) for \(k \neq j\) all vanish by symmetry and a relative \(N^{2D}\) is extracted from the quadratic terms after integration. The result is that

\[
\Delta I = \frac{1}{24N_{cell}^{2D}} \sum_{i=1}^{N_{cell}^{2D}} \sum_{j=1}^{N_{cell}^{2D}} \cdots \sum_{k=1}^{N_{cell}^{2D}} (u_{d,max} - u_{d,min})^2 \frac{\partial}{\partial u_i^2} H(u_i, u_k, \cdots u_{i,m}) \frac{\partial}{\partial u_j^2} \frac{\partial}{\partial u_k^2} \cdots \frac{\partial}{\partial u_{i,m}^2} H(u_i, u_k, \cdots u_{i,m}).
\]

(A.14)

Note that the one- and two-dimensional results can be obtained from the above equation. The critical feature to note is the overall \(N_{cell}^{2D}\) convergence rate. The more dimensions in the problem, the slower the convergence for numerical quadrature.

### A.3 Convergence of Monte Carlo Solutions

An alternative approach to solving Equation A.1 is the Monte Carlo method whereby \(N_{hist}\) particle histories are simulated. In this case, the Monte Carlo method converges to the true answer according to the central limit theorem (Feller, 1967) which is expressed as

\[
\frac{\Delta T_{MC}(x)}{T_{MC}(x)} = \frac{1}{\sqrt{N_{hist}}} \frac{\sigma_{MC}(x)}{T_{MC}(x)},
\]

where \(T_{MC}(x)\) is the tally calculated in a voxel located at \(x\) as calculated by the Monte Carlo method and \(\sigma_{MC}(x)\) is the variance associated with the distribution of \(T_{MC}(x)\). Note that this variance \(\sigma_{MC}(x)\) is an intrinsic feature of how the particle trajectories deposit energy in the spatial voxel. It is a “constant” for a given set of initial conditions and is conventionally estimated from the sample variance. It is also assumed, for the purpose of this discussion, that the sample variance exists and is finite.

### A.4 Comparison between Monte Carlo and Numerical Quadrature

The deterministic models considered in this discussion pre-calculate \(F(x - x', p')\) of Equation A.7 or \(F(x, x', p, p', \Delta x)\) of Equation A.10 storing them in arrays for iterative use. Then, during the iterative calculation phase, a granulated matrix operation is performed. The associated matrix product is mathematically similar to the “mid-point” \(N_s + N_p\)-multidimensional integration discussed previously:

\[
T(x) = \int_{F} du H(u, x),
\]

(A.16)

where \(D = N_s + N_p\) and \(u = (x_1, x_2, \cdots x_{N_s}, p_1, p_2, \cdots p_{N_p})\). That is, \(u\) is a multidimensional variable that encompasses both space and momentum. In the case of photon convolution, \(H(u, x)\) can be inferred from Equation A.6 and takes the explicit form:

\[
H(u, x) = \int dp F(|x - x'|, p')Q(x', p').
\]

(A.17)

There is a similar expression for the phase space evolution model.

The “mid-point” integration represents a “first-order” deterministic technique and is applied more generally than the convolution or phase space evolution applications. As shown previously, the convergence of this technique obeys the relationship:

\[
\frac{\Delta T_{NMC}(x)}{T_{NMC}(x)} = \frac{1}{N_{cell}^{2D}} \frac{\sigma_{NMC}(x)}{T_{NMC}(x)},
\]

(A.18)

where \(T_{NMC}(x)\) is the tally in a spatial voxel in an arbitrary \(N_s\)-dimensional geometry calculated by a non-Monte Carlo method where \(N_p\) momentum components are considered. The \(D\)-dimensional phase space has been divided into \(N_{cell}\) “cells” equally divided among all the dimensions so that the “mesh-size” of each phase space dimension is \(N_{cell}^{2D}\). The constant of proportionality as derived previously is

\[
\sigma_{NMC}(x) = \frac{1}{24} \sum_{i=1}^{N_{cell}^{2D}} \sum_{j=1}^{N_{cell}^{2D}} \cdots \sum_{l=1}^{N_{cell}^{2D}} (u_{d,max} - u_{d,min})^2 \frac{\partial^2}{\partial u_i^2} \frac{\partial^2}{\partial u_j^2} \cdots \frac{\partial^2}{\partial u_{l,m}^2} H(u_i, u_j, \cdots u_{l,m}),
\]

(A.19)

where the \(u\)-space of \(H(u)\) has been partitioned in the same manner as the phase space described above. \(u_{d,min}\) is the minimum value of \(u\) while \(u_{d,max}\) is its maximum value. \(u\) is the midpoint of the cell in the \(j\)th dimension at the \(j\)th mesh index.

The equation for the proportionality factor is quite complicated. However, the important point to note is, that it depends only on the second derivatives of \(H(u)\) with respect to the phase-space variables, \(u\). Moreover, the non-Monte Carlo proportionality factor is quite different from the Monte Carlo proportionality factor. It would be difficult to predict which would be smaller, and almost certainly, would be application dependent.

We now assume that the computation time in either case is proportional to \(N_{hist}\) or cell. That is, \(T_{MC} = \alpha_{MC}N_{hist}\) and \(T_{NMC} = \alpha_{NMC}N_{cell}\). In the Monte Carlo case, the computation time is simply \(N_{hist}\) times the average computation time/history. In the non-Monte Carlo case, the matrix operation can potentially attempt to connect every cell in the \(D\)-dimensional phase space to the tally at point \(x\). Thus, a certain number of floating-point and integer operations are required for each cell in the problem.
Consider the convergence of the Monte Carlo and non-Monte Carlo method. Using the above relationships, one can show that:

$$\frac{\Delta T_{MC}(x)/T_{MC}(x)}{\Delta T_{NMC}(x)/T_{NMC}(x)} = \left( \frac{\sigma_{NMC}(x)}{\sigma_{MC}(x)} \right)^{1/2} \frac{\sigma_{NMC}^\alpha}{\sigma_{MC}^\alpha} t^{(4-D)/2}, \quad (A.20)$$

where $t$ is the time measuring computational effort for either method. We have assumed that the two calculation techniques are the same. Therefore, given enough time, $D_{MC}(x) = D_{NMC}(x)$. One sees that, given long enough, the Monte Carlo method is always more advantageous for $D > 4$. We also note, that inefficient programming in the non-Monte Carlo method is severely penalized in this comparison of the two methods.

Assume that one desires to do a calculation to a prescribed $\epsilon = \Delta T(x)/T(x)$. Using the relations derived so far, we calculate the relative amount of time to execute the task to be:

$$\frac{t_{NMC}}{t_{MC}} = \left( \frac{\sigma_{MC}}{\sigma_{NMC}} \right)^{D/2} \frac{T_{NMC}(x)/\sigma_{NMC}(x)}{T_{MC}(x)/\sigma_{MC}(x)} e^{(4-D)/2}, \quad (A.21)$$

which again shows an advantage for the Monte Carlo method for $D > 4$. Of course, this conclusion depends somewhat upon assumptions of the efficiency ratio $\sigma_{MC}/\sigma_{NMC}$ which would be dependent on the details of the calculation technique. Our conclusion is also dependent on the ratio $[\sigma_{NMC}(x)/T_{NMC}(x)]^{D/2} / [\sigma_{MC}(x)/T_{MC}(x)]$ which relates to the detailed shape of the response functions. For distributions that can vary rapidly, the Monte Carlo method is bound to be favored. When the distributions are flat, non-Monte Carlo techniques may be favored.

Nonetheless, at some level of complexity (large number of $N_{cal}$'s required) Monte Carlo becomes more advantageous. Whether or not one's application crosses this complexity “threshold,” has to be determined on a case-by-case basis.

Smaller dimensional problems will favor the use of non-Monte Carlo techniques. The degree of the advantage will depend on the details of the application.

References


Alsmiller R. G., Jr and Moran H. S. The electron–photon cascade induced in lead by photons in the energy range 15 to 100 MeV. Report ORNL-4192, Oak Ridge National Laboratory, Oak Ridge, Tennessee, 1968.


Nicoli D. F. The application of Monte Carlo cascade shower generation in lead. Submitted in partial fulfillment of the requirement for the degree of Bachelor of Science at the Massachusetts Institute of Technology, 1966.


2 Basics of Monte Carlo Simulations

2.1 Monte Carlo Method

Monte Carlo (MC) techniques are widely used in natural and social sciences. There are many different “flavors” of how to work with these techniques. As we have seen in the previous chapter, there is a long tradition of using MC methods in different areas, including medical physics. Therefore, it is difficult to provide a general definition of the MC method. Consequently, a literature search results in many definitions. Some examples of introducing literature, useful to read before working with MC methods, are the corresponding textbook chapters in Numerical Recipes in C (Press et al., 1992) or in “The Review of Particle Physics” (Nakamura et al., 2010) of the Particle Data Group. A nice introduction is also provided by James in “Monte Carlo Theory and Practice” (James, 1980). More references on MC techniques and random number sampling can be found in these reviews.

For our purposes, we define the MC method in the following short way:

Monte Carlo is a numerical method to solve equations or to calculate integrals based on random number sampling.

The two aspects of this definition, random number sampling and numerical integration, are outlined in detail in the next two sections.

2.1.1 Random Number Sampling

MC algorithms use a computer program, a procedure or a subroutine, called “random number generator” (RNG). However, computers cannot really generate “random” numbers because the output of any program is—by definition—predictable; hence, it is not truly “random.” Therefore, the result of these generators shall correctly be termed “pseudorandom numbers.”

A huge sequence of these pseudorandom numbers is required to solve a complex problem. The numbers within a sequence of random numbers shall be uncorrelated, that is, they must not depend on each other. Because this is impossible with a computer program, they should at least appear independent. In other words, any statistical test program shall show that the numbers within the sequence are uncorrelated and any computer code that requires independent random numbers shall produce the same result with different sequences. If this is the case within the uncertainty of the simulation, then these sequences can be called pseudorandom. To keep the notation short throughout this book, we will nevertheless call them just “random numbers.” However, we have to keep in mind the real character of these numbers.

A pseudo-RNG must be examined carefully before it can be used for a specific purpose. A useful generator for simulations in radiation therapy must provide two important features:

- The period of the sequence shall be large enough. Otherwise, if the sequence is reused several times, the results of the MC simulation are correlated.
- They must be uniformly distributed in multiple dimensions. That means random vectors created from an n-tuple of random numbers must be uniformly distributed in the n-dimensional space. Typically, it is not obvious how to detect correlations in higher dimensions.

Most of the generators produce uniformly distributed random numbers in some interval, typically in [0,1]. It is useful to have a look at simple RNGs to understand their operating principle. One class of simple RNGs is called linear congruential generators. They generate a sequence of integers \( I_1, I_2, I_3, \ldots \), each between 0 and \( m - 1 \) by the recurrence relation:

\[
I_{j+1} = aI_j + c \pmod{m}
\]

with the parameters

\[ a : \text{ multiplier} \]
\[ c : \text{ increment} \]
\[ m : \text{ modulus} \]
An example is the quick and dirty generator used by old EGS4 (Nelson et al., 1985) implementations:

\[ I_{j+1} = a I_j \]  

(2.2)

with

\[ a = 663608941 \]

\[ c = 0 \]

\[ m = 2^{32} \]

On a machine with 32-bit integer representation, the product of two unsigned integers is the low-order 32 bits of the true 64-bit result. Because of \( m = 2^{32} \), the modulus is taken into account automatically. Therefore, this RNG is very fast. However, its sequence length of \( 2^{32} \) is not enough for MC applications in radiotherapy.

Computer operating system-supplied RNGs are typically of the same linear congruential type. Hence, it is not recommended to trust them. For MC applications in radiotherapy, long-sequence RNGs are required. They also should be portable, that is, the same sequence should be produced on different machines. A good source of high-quality RNGs including test programs to examine the created sequences in multiple dimensions is the CERN program library (http://www.cern.ch).

A class of long-sequence RNGs called “subtract-with-borrow” algorithms has been developed by Marsaglia and Zaman (1991). The CERN library function RANMAR for instance has a sequence length of \( 2^{144} \). It is used, for example, by EGS4 (Nelson et al. 1985) and XVMC (Fippel, 1999).

In EGSnrc (Kawrakow, 2000a), the use of RANLUX (Lüscher, 1994) is recommended. It allows different luxury levels between 0 and 4. The quality and also the simulation time increase with increasing luxury level. RANMAR can also be selected in EGSnrc. According to the CERN library documentation, its quality corresponds to RANLUX with luxury level between 1 and 2. According to the EGSnrc manual (Kawrakow et al., 2011), with luxury level 1 or higher, no problems have been discovered in practical EGSnrc calculations.

### 2.1.2 Numerical Integration

Function \( y = f(x) \) shall be integrated in interval \([a,b]\), that is, the area \( A \) enclosed by function \( f(x) \), the \( x \)-axis as well as the interval limits \( a \) and \( b \) shall be calculated (see Figure 2.1):

\[ A = \int_a^b f(x) \, dx. \]  

(2.3)

If this is impossible analytically, some numerical method must be applied. One of the many numerical options is called MC integration because it is based on a sequence of uniformly distributed random numbers. Generally, a computer-generated random number \( \eta_i \) is uniformly distributed in interval \([0,1]\). It can be scaled to interval \([a,b]\) by

\[ \xi_i = (b-a) \eta_i + a, \]  

(2.4)

that is, \( \xi_i \) is uniformly distributed in \([a,b]\). Some first rough estimate of the real area \( A \) is now given by (see Figure 2.1)

\[ A_1 = (b-a) f(\xi_1), \]  

(2.5)

that is, the rectangle given by the function value at random point \( \xi_1 \) and the length of the interval. This estimate is really rough; therefore, we should do this a second time:

\[ A_2 = \frac{1}{2} [(b-a)f(\xi_2) + A_1] = \frac{b-a}{2} \{ f(\xi_1) + f(\xi_2) \}. \]  

(2.6)

Note that we have averaged the areas from both runs providing in this way a better estimate of the real integral. The generalization is now obvious; after \( N \) runs (with \( N \) random numbers), we obtain

\[ A_N = \frac{b-a}{N} \sum_{i=1}^{N} f(\xi_i) = (b-a) \langle f(x) \rangle \]  

(2.7)

with the average function value for \( N \) samples:

\[ \langle f(x) \rangle \equiv \frac{1}{N} \sum_{i=1}^{N} f(\xi_i). \]  

(2.8)

The basic theorem of MC integration (Press et al., 1992) also provides information on the uncertainty of the estimate

\[ A = A_N \pm (b-a) \sqrt{ \frac{\langle f^2(x) \rangle - \langle f(x) \rangle^2 \rangle}{N} }. \]  

(2.9)

The average of the function value squared is defined by

\[ \langle f^2(x) \rangle \equiv \frac{1}{N} \sum_{i=1}^{N} f^2(\xi_i). \]  

(2.10)
The estimated area $A_N$ converges to the real integral $A$ in the limit $N \to \infty$. The convergence is slow because of the $1/\sqrt{N}$ behavior, that is, the statistical uncertainty is reduced by a factor of 2 if the number of random points $N$ (and the calculation time) is increased by a factor of 4. Therefore, the MC method should not be used for this simple type of numerical integration. There are better options; see, for example, Press et al. (1992). However, the MC method comes into play if all other methods fail, for example, if the dimensionality of the problem becomes very large, that is, the integral shall be calculated in a space with 10, 100, or even an infinite number of dimensions.

Let us assume that function $f(\vec{x})$ shall be integrated in volume $V$ of a space with $D$ dimensions. Instead of random numbers for MC integration, random points (or vectors) uniformly distributed in the multidimensional volume $V$ are required. Because volume $V$ has $D$ dimensions, we need $D$ random numbers to form one random point. To sample $N$ random points in this volume, we need $D \times N$ random numbers. Therefore, the quality of the RNG should be checked for higher dimensions, not just for the 1D, 2D, or 3D case. Often, correlations can be observed in higher dimensions even if the RNG looks uniform in lower dimensions.

Now, let us assume that we have sampled $N$ random points $\vec{\xi}_1, \ldots, \vec{\xi}_N$ uniformly distributed in multidimensional volume $V$. Then the basic theorem of MC integration is given by (Press et al., 1992)

$$\int \! dV \, f(\vec{x}) = V \langle f(\vec{x}) \rangle \pm \sqrt{\frac{\langle f^2(\vec{x}) \rangle - \langle f(\vec{x}) \rangle^2}{N}}$$

(2.11)

with

$$\langle f(\vec{x}) \rangle = \frac{1}{N} \sum_{i=1}^{N} f(\vec{\xi}_i) \quad \text{and} \quad \langle f^2(\vec{x}) \rangle = \frac{1}{N} \sum_{i=1}^{N} f^2(\vec{\xi}_i).$$

(2.12)

Multidimensional numerical integration is necessary to solve the system of coupled transport equations for problems in radiation therapy, for example, for dose calculation. It is a system of equations because the transport problem for photons and electrons (positrons) must be solved. The system is coupled because electrons influence the photon transport (bremsstrahlung) and vice versa (Compton scatter, photon electric absorption, pair production). Theoretically, the problem has an infinite parameter space because the number of secondary photons and electrons is physically unlimited when we start with a primary particle of definite energy. Therefore, the numerical integration has to be performed in a space with infinite dimensions. Practically, the dimensionality is limited because the region of interest is limited and we usually stop the simulation if the photon or electron energy falls below some minimum energy.

Nevertheless, for numerical integration using MC, a random point in a high-dimensional parameter space must be sampled. This point can be demonstrated by a so-called “particle history,” a shower of secondary particles generated by a primary particle including all daughter particles. An example of a particle history is shown in Figure 2.2. This shows schematically what happens during an MC radiation transport simulation. It starts with a primary particle emitted from a particle source. The simulation takes into account the geometry of the problem, for example, the linac head geometry given by the technical specifications and/or the patient anatomy given by CT images. It also takes into account the material transport properties given by cross-section data. For the primary photon in the example of Figure 2.2, the distance to the first interaction site must be sampled based on the total cross section of the corresponding medium. At the interaction site, the type of interaction (Compton scatter in this case) must be sampled. All secondary particle parameters (secondary particle energies and scattering angles) are determined using differential cross sections and the corresponding probability distribution functions. These steps are repeated until the primary and secondary particles have left the simulation geometry or the particle energy falls below some minimum energy. Dose, for example, is calculated by accumulating the absorbed energy per region. It is obvious that a huge number of particle histories has to be simulated to obtain a result of low noise. Because of Equation 2.11, the noise level and the statistical uncertainty can be reduced by a factor of 2 if the number of histories is increased by a factor of 4. This causes very long calculation times for transport simulation problems in radiation therapy.

2.1.3 Nonuniform Sampling Methods

As we have learned in Section 2.1.1, RNGs usually generate uniformly distributed random numbers. However, for MC transport simulations in radiation therapy, random numbers distributed according to specific probability weight distribution functions $p(x)$ are required many times within the algorithm.

Let us assume that we want to generate a random number $\xi$ in interval $[a, b]$ and distributed according to the nonuniform probability weight function $p(x)$. Available is a RNG generating only uniformly distributed random numbers $\eta$ in interval $[0, 1]$
[0,1]. How can $\xi$ be sampled from $p(x)$? This is possible using the cumulative distribution function $P(x)$ defined by the integral

$$P(x) = \int_a^x p(x') \, dx', \quad a \leq x \leq b, \quad P(a) = 0, \quad P(b) = 1. \quad (2.13)$$

Function $P(x)$ is monotonically increasing in interval $[a,b]$. The function values $y = P(x)$ are limited to interval $[0,1]$. If formula $y = P(x)$ can be transposed with respect to $x$, that is, the inverse can be determined as

$$x = P^{-1}(y), \quad (2.14)$$

then it can be shown that

$$\xi = P^{-1}(\eta), \quad (2.15)$$

is distributed according to probability weight function $p(x)$. For a proof, see, for example, Nelson et al. (1985). This method of nonuniform random number sampling is called the direct or transformation method. It depends on the fact that the inverse of $P(x)$ can be easily and efficiently calculated.

If this is impossible, the indirect or rejection method (Nelson et al. 1985) has to be applied instead. It takes into account an adequately chosen comparison function $g(x)$ with $g(x) > p(x)$ in interval $[a,b]$. A further condition of $g(x)$ is that a random number $\xi_0$ can be sampled easily from $g(x)$, for example, using the transformation method. Now, we sample a new uniform random number $\eta_0$ from interval $[0,g(\xi_0)]$. We accept $\xi_0$ as valid random number if $\eta_0 < p(\xi_0)$. We reject $\xi_0$ if $p(\xi_0) \leq \eta_0 \leq g(\xi_0)$. In case of rejection, we start with sampling a new random number $\xi_0$ from $g(x)$ and so on. Again, we do not prove here that $\xi_0$ sampled in this way is distributed according to $p(x)$. But the rejection method can be demonstrated using the simple example of a constant rejection function $g(x) = p(x_{\text{max}}) = \text{const}$, with $x_{\text{max}}$ as the position of the maximum of $p(x)$ in interval $[a,b]$. In this case, $\xi_0$ is just sampled from a uniform distribution in $[a,b]$. With the rejection random number $\eta_0 < p(\xi_0)$, it is obvious that $\xi_0$ is distributed according to $p(x)$.

Of course, the rejection method should be used with care. That is, comparison function $g(x)$ should not differ too much from $p(x)$. If the deviation is too large, the number of rejections can become too large and the method becomes inefficient.

2.2 Monte Carlo Transport in Radiation Therapy

The basic concepts to simulate the transport of photons, electrons, positrons, neutrons, protons, or heavy ions in the energy range of radiation therapy are complex. Therefore, this chapter can only provide the fundamentals of the methodology. To understand all techniques in full detail, the reader is referred to corresponding literature, like the textbook Monte Carlo Transport of Electrons and photons (Jenkins et al., 1988). A lot of information is available also in the user manuals of frequently used MC packages, like EGS4 (Nelson et al., 1985), EGSnrc (Kawrakow, 2000a; Kawrakow et al., 2011), Penelope (Salvat et al., 2011), MCNP (Briesmeister, 1997), and GEANT4 (Agostinelli et al., 2003). More basic literature is listed in review papers about MC treatment planning (Chetty et al., 2007; Reynaert et al., 2007).

In the next sections, we show how particle transport collision-by-collision is simulated in detail. This so-called analog particle transport simulation scheme is demonstrated using the photon as an example. A further section introduces the condensed history (CH) technique that allows efficient simulations of charged particle transport in radiotherapy.

2.2.1 Analog Particle Transport

Let us assume that a photon of energy $E$ hits the surface of a homogeneous medium. Then, the probability $p(s)$ that this photon interacts after path length $s$ with the medium is given by the attenuation law:

$$p(s) \, ds = \mu(E) \, e^{-\mu(E)s} \, ds. \quad (2.16)$$

The parameter $\mu(E)$ is called the linear attenuation coefficient of the medium for photons of energy $E$. The mean free path length $\langle s \rangle$ until interaction can be calculated from this distribution function if the medium is extended infinitely below the surface:

$$\langle s \rangle = \int_0^\infty ds \, p(s) = \mu(E) \int_0^\infty ds \, e^{-\mu(E)s} = \frac{1}{\mu(E)}. \quad (2.17)$$

This allows us to express the attenuation law (2.16) in terms of the number of mean free path lengths:

$$\lambda = \frac{s}{\langle s \rangle} = \mu(E) \, s, \quad (2.18)$$

that is

$$p(\lambda) \, d\lambda = e^{-\lambda} \, d\lambda. \quad (2.19)$$

The advantage of this notation is that it also works for heterogeneous geometries if the number of mean free path lengths is defined by

$$\lambda = \sum_{i=1}^n \mu_i(E) \, s_i. \quad (2.20)$$

To calculate $\lambda$, the photon must be traced on a straight line from the Start position on the surface through different regions $i$ containing different materials until the interaction point $P$. In
of Monte Carlo Simulations

2.2.2 Charged Particle Transport

2.2.2.1 Condensed History Technique

Section 2.2.1 outlined the fundamental procedure to simulate the transport of any particle type through matter. In general, this is the standard simulation method for neutral particles because the free path length between two interactions is in the order of the size of the simulation geometry. For example, the mean free path lengths of photons in the therapeutic energy range are in the order of 10 cm in water and human tissue. The region of interest for dose calculation in radiation therapy has a size of about 30 cm. Therefore, on average, very few photon interactions are simulated.

This is completely different for charged particles like electrons or protons. For radiation therapy energies, they undergo a very large number of single interactions. Consequently, the simulation of one electron (positron or proton) history would require a much longer calculation time than the simulation of one photon history. Hence, this approach is impractical for most of the transport problems in radiation therapy.

Fortunately, almost all of these interactions are elastic or semielastic. That means no energy or a small amount of energy is transferred from the charged particle to the surrounding matter. Furthermore, the particle direction changes in general only by small scattering angles. This allows us to group many of these elastic and semielastic events into one CH step. The method has been introduced in 1963 by Berger (1963) and is called CH technique. The majority of MC algorithms, applied in radiotherapy, perform electron, positron, proton, or heavier charged particle transport using the CH technique.

Present-day CH implementations divide all interactions of one charged particle history into hard and soft collisions as well as hard and soft bremsstrahlung production events. The two collision types are distinguished by an arbitrary kinetic energy

and scattering angles, are sampled using the probability distributions given by the corresponding differential cross sections. Furthermore, kinematic conservation laws must be taken into account. In general, for this purpose, the transformation method does not work, that is, the rejection technique is the method of choice.

After that, everything is known to repeat the three steps with the secondary particles. Even electrons and positrons could be simulated in this analog manner. The whole particle history (see an example in Figure 2.2) is simulated including all secondary particles and its daughter particles. The transport simulation of a particle stops if it leaves the geometry of interest or its energy falls below some predefined minimum energy. These cut-off parameters are usually denoted as $P_{\text{cut}}$ for photons or $E_{\text{cut}}$ for charged particles. During each step of the history, the values of interest are calculated for accumulation. For example, the absorbed energy per region (voxel) is accumulated if the dose is calculated. According to Equation 2.11, the number of simulated particle histories determines the statistical accuracy and thus the calculation time.
loss threshold $E_c$. Hard and soft bremsstrahlung production is distinguished using the parameter $k_c$. Collision events with an energy transfer lower than $E_c$ to secondary electrons are called soft collisions. These soft collisions are simulated implicitly by continuous energy transfer from the charged particle onto the matter surrounding the particle track. The direction change of the particle due to many small angles is simulated by one large multiple scattering angle. All hard collisions are simulated explicitly as in the case of photons. The minimum energy of secondary particles created during hard collisions is equivalent to $E_c$. It also provides the maximum energy and consequently the maximum range of charged secondary particles produced during soft collisions. This range has to be smaller than the spatial resolution of the simulation geometry. The meaning of the bremsstrahlung production threshold $k_c$ is similar. Therefore, the arbitrary parameters $E_c$ and $k_c$ must be chosen with care. The simulation result can be influenced negatively if they are too large. On the other hand, the simulation can last too long if these parameters are chosen too small. Please note that the parameter $E_c$ must not be confused with the particle track end energy $E_{cut}$. However, in many MC simulations, both parameters are chosen to be equal.

Because of the approximate nature of the CH transport, it is furthermore useful to limit the maximum distance traveled in one CH step. This could be realized by another arbitrary user parameter, the global maximum step size $s_{\text{max}}$ or by material- and mass density-dependent parameters $s_{\text{max}}^{\text{mat}}$. In many MC algorithms, the maximum step size is determined based on the percentage maximum energy loss $E_{\text{step}}$. This way, the maximum spatial step size automatically depends on the stopping power and the mass density of the present material.

With all of these possibilities, the end of one CH step is determined either by the maximum step size or by the next hard interaction. As a result, a complete electron history may look like the example shown in Figure 2.3. Electrons move in general on straight lines during the CH step. They change the direction due to multiple scattering at the end of the step (see Figure 2.4) or (as in the case of Figures 2.3 and 2.5) between step limits. In the second case, the position of the multiple scattering can be determined, for example, randomly using the so-called random hinge method. Figure 2.3 shows that the first CH step is limited by a hard Møller interaction. Møller interactions result in secondary electrons, also called delta electrons. The delta electrons are simulated in the same way using the CH technique as in the case of photons using the analog technique. As a result of the CH transport, the simulated path length must be corrected and a transversal displacement has to be taken into account.

The various components of the CH technique are outlined in the following subsections using the electron as an example.

### 2.2.2.2 Continuous Energy Loss

During a CH step the charged particle continuously loses energy due to soft interactions. The average energy loss $dE$ per CH step length $ds$ at point $\vec{r}$ is given by the restricted linear stopping power:

$$L(\vec{r}, E, E_c, k_c) \equiv -\frac{dE}{ds}_{\text{res}} = L_{\text{coll}}(\vec{r}, E, E_c) + L_{\text{rad}}(\vec{r}, E, k_c)$$  \hspace{1cm} (2.26)$$

with

$$L_{\text{coll}}(\vec{r}, E, E_c) \equiv \left( \frac{dE}{ds} \right)_{\text{coll}} \quad \text{and} \quad L_{\text{rad}}(\vec{r}, E, k_c) \equiv \left( \frac{dE}{ds} \right)_{\text{rad}}$$  \hspace{1cm} (2.27)$$

$$\text{FIGURE 2.3} \quad \text{Example of a particle history starting with a primary electron } e^- \text{ (dashed line) via multiple scatter, Møller interactions, and bremsstrahlung production events leading to secondary (delta) electrons } e^- \text{ (dashed lines) and secondary photons } p \text{ (straight lines).}$$

$$\text{FIGURE 2.4} \quad \text{Example of a simulated electron path using the CH technique and multiple scatter compared to a possible real electron path (or simulated using the analog technique). As a result of the CH transport, the simulated path length must be corrected and a transversal displacement has to be taken into account.}$$

$$\text{FIGURE 2.5} \quad \text{Same as Figure 2.4, but this time the electron is simulated using the random hinge method. This way, path length corrections and transversal displacements are approximately taken into account.}$$
defined as restricted linear collision and radiation stopping powers. They can be calculated using the collision cross section \( \sigma_{\text{col}}(\vec{r}, E, E') \) and the bremsstrahlung production cross section \( \sigma_{\text{rad}}(\vec{r}, E, E') \) by

\[
\begin{align*}
L_{\text{col}}(\vec{r}, E, E_s) &= N(\vec{r}) \int_0^{E_s} dE' \sigma_{\text{col}}(\vec{r}, E, E') \\
L_{\text{rad}}(\vec{r}, E, k_s) &= N(\vec{r}) \int_0^{k_s} dk' \sigma_{\text{rad}}(\vec{r}, E, k').
\end{align*}
\] (2.28)

\( N(\vec{r}) \) is the number of scattering targets per unit volume at point \( \vec{r} \). Note the stopping power integration is restricted to energies below \( E_s \) and \( k_s \). That means the energy transfer to secondary charged particles is restricted to be below \( E_s \) and the energy transfer to secondary photons is restricted to be below \( k_s \). Interactions with higher energy transfer are simulated explicitly within this CH scheme.

The step length \( s \) for an electron with initial energy \( E_0 \) that loses energy \( \Delta E \) due to CH transport can be calculated by integrating Equation 2.26:

\[
s = -\int_0^{E_0} \frac{dE}{\sigma_{\text{col}}(\vec{r}, E, E_s, k_s)} = \int_0^{E_1} \frac{dE}{\sigma_{\text{col}}(\vec{r}, E, E_s, k_s)}
\] (2.29)

with \( E_1 = E_0 - \Delta E \) being the electron energy at the end of the step. The minus sign in the definition (2.26) is used here to exchange the upper and lower integration limits. The integration (2.29) is performed along the simulated electron path and takes into account the spatial dependency of the stopping power on \( \vec{r} \). That means function \( L(\vec{r}, E, E_s, k_s) \) changes accordingly if a boundary to a region with different material is crossed.

Equation 2.29 provides a unique functional dependency of the CH step length \( s \) and the step end energy \( E_1 \). Therefore, all electrons with initial energy \( E_0 \) and transported along the same step will reach the step end with the same energy \( E_1 \). Here, it becomes clear that this approach is an approximation because in reality, only the mean step end energy can be calculated by Equation 2.29. Analog MC simulations show that the real step end energies are random according to a distribution with a mean value of \( E_1 \). This effect is known as energy straggling. In the CH simulation scheme defined here, we have to distinguish between soft energy straggling and hard energy straggling. Hard energy straggling is simulated explicitly and the corresponding effects are correctly taken into account. Soft or subthreshold energy straggling is either neglected or it is simulated by an adequate straggling distribution function.

Indeed, subthreshold straggling can be neglected during a CH MC simulation of specific problems if we choose the parameter \( E_1 \) and the step size small enough. In this case, energy straggling is dominated by the explicitly modeled hard ionization events with energy transfer larger than \( E_1 \) and the low energy fluctuations have a negligible influence on the final result. Furthermore, in many cases, subthreshold energy straggling is additionally insignificant compared to the influence of the range straggling (see Section 2.2.2.4).

A potential disadvantage of neglecting soft energy fluctuations is that the calculation time can increase unnecessarily. Therefore, the implementation of some subthreshold straggling theory can help to speed up the simulation or to avoid artifacts. For example, a Gaussian distribution function according to an approach by Bohr (1948) can be used to model energy straggling. More accurate are the theories of Landau (1944) and Vavilov (1957).

### 2.2.2.3 Multiple Scattering

In contrast to reality and as demonstrated in Figure 2.4, in a CH approach, charged particles move on straight lines during one step. The combined effect of many small-angle elastic and semielastic collisions during one step is simulated by sampling the angular deflection based on a dedicated multiple scattering theory. An example is the theory developed by Fermi and Eyges (Eyges, 1948). This theory models the probability \( p(\theta, \varphi) d\theta d\varphi \) that the electron is scattered within the solid multiple scattering angular section \([\theta + d\theta, \varphi + d\varphi] \) as 2D Gaussian distribution:

\[
p(\theta, \varphi) d\theta d\varphi = \frac{\theta}{\pi} e^{-\frac{\theta^2}{\theta^2(s)}} d\theta d\varphi,
\] (2.30)

where \( \theta \) is the azimuthal multiple scattering angle, \( \varphi \) is the polar multiple scattering angle, and \( \theta^2(s) \) is the mean square deflection angle after step length \( s \). The distribution (2.30) is not normalized because \( \theta \) is limited to the interval \([0, \pi]\) instead of \([0, \infty]\).

Equation 2.30 results in two separate cumulative distribution functions:

\[
P_{\theta}(\theta) = 1 - \exp\left(-\frac{\theta^2}{\theta^2(s)}\right), \quad P_{\varphi}(\varphi) = \frac{\varphi}{2\pi}.
\] (2.31)

The transformation method from Section 2.1.3 and a uniform random number \( \xi_\theta \) can be applied to sample \( \theta \):

\[
\theta = \sqrt{-\theta^2(s)\ln(1 - \xi_\theta)}.
\] (2.32)

It has to be noted that \( \xi_\theta \) is uniformly distributed in an interval \([0, \xi_{\theta_{\text{max}}}]\) with \( \xi_{\theta_{\text{max}}} < 1 \) to ensure \( \theta \leq \pi \). Another option is to sample \( \xi_\varphi \) from \([0,1]\) and to reject \( \theta \) if \( \theta > \pi \). Because of the rotational symmetry, the polar scattering angle \( \varphi \) is determined from a uniform distribution in \([0,2\pi]\).

The quantity \( \theta^2(s) \) is calculated using the linear scattering power \( T_s(\vec{r}, E) \) at point \( \vec{r} \):

\[
\theta^2(s) = \int_0^s ds' T_s(s', E).
\] (2.33)
The material parameter $T_i(\vec{r}, E)$ depends on the atomic composition at point $\vec{r}$ but also on the electron energy $E$. This must be taken into account in Equation 2.33 because the electron loses energy between the beginning and the end of the step.

The Gaussian (2.30) is a good approximation of the reality for small cumulative scattering angles $\theta$. However, large scattering angles are underestimated using this distribution. Therefore, very often in the past (e.g., in EGS4 (Nelson et al., 1985)) the multiple scattering distribution of Molière (1948) was used. Nevertheless, this improvement is also based on the small-angle approximation, and especially for large angles, the Molière theory still has limitations.

Most suitable for MC simulations are algorithms based on the exact theory of Goudsmit and Saunderson (1940a,b). Examples are the multiple scattering algorithms as implemented in MCNP (Briesmeister, 1997), Penelope (Salvat et al., 2011), and EGSnrc (Kawrakow et al., 2011). These algorithms sample multiple scattering angles close to reality even for large angles.

### 2.2.2.4 Transport Mechanics

Figure 2.4 compares a CH electron step to some possible real electron path or to an electron path simulated using analog MC transport. The figure shows that several problems can be expected from a CH simulation. For better comprehension, in this section, we neglect energy loss during the CH step. That means here the electron energy has not changed between the beginning and the end of the step and all interactions during the step are assumed to be elastic. As shown in Figure 2.4, many MC codes sample the multiple scattering angle at the end of the step, that is, the electron moves on a straight line until the final position and then it changes the direction due to multiple scatter. It is obvious that in this case, the electron distance is overestimated. Also shown in Figure 2.4, the real electron path is curved. Therefore, the real electron range is shorter if we assume that both electrons (the real electron and the CH electron) move with the same path length. Furthermore, the real electron range fluctuates around some mean value. This effect is called range straggling and should not be confused with energy straggling as discussed in Section 2.2.2.2. Range straggling is independent of the energy loss of the electron.

Many CH history algorithms employ a path length correction (PLC) algorithm to take into account range overestimation and range straggling. Besides this, a transverse displacement (TD) algorithm is useful to take into account transverse fluctuations of the real electron end position relative to the lateral position as simulated during the CH step. A very simple PLC and TD approach, called random hinge method, is demonstrated in Figure 2.5. It is implemented, for example, in Penelope (Salvat et al., 2011) or XVMC (Kawrakow and Fippel, 2000). In this method, the whole step length $s$ is subdivided using a uniformly distributed random number $\xi$ from interval $[0,1]$ into two substeps $\xi s$ and $(1 - \xi)s$. The multiple scattering angle is sampled between the two substeps instead at the end of the step. The average longitudinal (PLC) and lateral (TD) displacements simulated in this manner are very close to the exact values. More accurate is the parameter-reduced electron-step transport algorithm developed by Kawrakow and Bielajew (1998) (sometimes called PRESTA-II) as implemented in EGSnrc (Kawrakow, 2000a). This algorithm reproduces first- and second-order spatial moments to within 0.1%.

The problems and algorithms discussed in this section could be neglected completely if the charged particle transport steps are small enough. In the limit of infinitesimal small step sizes, any CH algorithm will converge to the correct answer. However, in this way, the simulation becomes extremely inefficient. Therefore, with the implementation of a sophisticated algorithm for the transport mechanics, charged particle transport can be used efficiently for MC simulations in radiation therapy.

#### 2.2.2.5 Boundary Crossing

Theoretically, charged particle MC transport using the CH technique with multiple scattering works only in homogeneous regions containing one definite medium. The problem of conditions with more than one material is shown schematically in Figure 2.6. It shows two regions consisting of different materials as well as two electron tracks, a real electron track calculated using the single scatter scheme and a CH simulated electron path. By definition, the CH technique shall simulate the condensed effect of a huge number of real electron paths, including the real path shown in Figure 2.6. This path is partly located in material I and has to take into account (for this part of the path) the interaction properties of material II. For the rest of the path, it has to consider the properties of material I. Other real electrons may be tracked completely in material I.

Figure 2.6 clearly demonstrates that, for arbitrarily shaped material interfaces, an exact theory of multiple scattering does not exist. Consequently, the CH technique can only be used in regions where the distance to the next material boundary is much larger than the size of the present electron step. Conversely, this means that the maximum step size $s_{\text{max}}$ must be decreased if an electron comes close to a material interface and it can be increased again if the electron moves away from the boundary. This kind of step size variation has been introduced, for example, in the original version of PRESTA (Bielajew and Rogers 1987). For geometries with many small regions of different material,
this leads to small step sizes and long calculation times. A further problem is that before an electron crosses the boundary, its CH step size becomes infinitesimally small. To permit boundary crossing anyway, the step size reduction must be switched off if a minimum step size is achieved. However, then the step size becomes similar to the boundary distance and step size artifacts may occur as we have seen in Figure 2.6 and the discussion above.

To avoid step size artifacts, a more accurate algorithm has been implemented in EGSnrc (Kawrakow, 2000a). Here, a minimum distance \(d_{\text{min}}\) to the next material interface can be defined by the user and CH transport is performed only if the electron is farther away from the boundary than this minimum distance. If the electron comes closer to the boundary than \(d_{\text{min}}\), the electron is transported in single scatter mode using the analog technique. This algorithm is free of step size artifacts. However, it also means that electrons are simulated in pure single scatter mode within geometries with very small regions, that is, if the average diameter of one region is smaller than \(d_{\text{min}}\). Consequently, calculation times can increase dramatically. Therefore, this technique including an appropriate choice of parameter \(d_{\text{min}}\) makes sense only for simulations with high accuracy requirements. An example is the simulation of ion chamber responses (Kawrakow, 2000b).

For the majority of MC simulations in radiation therapy, artifacts due to boundary crossing can be ignored because of their negligible influence on the result. A useful approach is, for example, to simulate electron boundary crossing without stopping at the interface using the random hinge technique. Here, the multiple scattering properties of material I are used if the “hinge” is sampled in material I; otherwise, the properties of material II are used. The total electron step length depends on the stopping powers and the corresponding step segments in both materials. According to the PENELOPE user guide (Salvat et al., 2011), this approach provides a “fairly accurate description of interface crossing.” It is, for example, an efficient and accurate method for linear accelerator head modeling or an MC-based dose calculation algorithm in radiation therapy treatment planning.

### 2.2.3 Cross Sections

#### 2.2.3.1 Photon Interaction Coefficients

In Section 2.2.1, the linear attenuation coefficient \(\mu(E)\) has been introduced to sample the distance between two photon interactions. For elements, compounds and mixtures, this material parameter can be calculated from the total cross sections \(\sigma_i(E)\) of the \(i\)-th element in the material as

\[
\mu(E) = \sum_i N_i(\bar{r}) \sigma_i(E).
\]  

(2.34)

\(N_i(\bar{r})\) is the number of atoms of element \(i\) per unit volume at point \(\bar{r}\) and is calculated by

\[
N_i(\bar{r}) = \frac{\rho(\bar{r}) w_i(\bar{r})}{m_i A_i(\bar{r})} = \frac{\rho(\bar{r}) N_A w_i(\bar{r})}{M_i(\bar{r})}.
\]  

(2.35)

\(\rho(\bar{r})\) is the mass density at point \(\bar{r}\), \(w_i(\bar{r})\) is the weight fraction of the \(i\)-th element at point \(\bar{r}\), \(A_i(\bar{r})\) is the relative atomic mass of the \(i\)-th element at point \(\bar{r}\), \(M_i(\bar{r})\) is molar mass of the \(i\)-th element at point \(\bar{r}\), \(m_i = 1.6605388 \cdot 10^{-27}\) kg is the atomic mass unit, and \(N_A = 6.0221418 \cdot 10^{23}\) mol\(^{-1}\) is the Avogadro constant (Nakamura et al. 2010).

Corresponding to Equation 2.23, the total atomic cross sections \(\sigma_i(E)\) are given by the sum of the cross sections for the different processes:

\[
\sigma_i(E) = \sigma_{i,A}(E) + \sigma_{i,R}(E) + \sigma_{i,C}(E) + \sigma_{i,P}(E),
\]  

(2.36)

that is, photoelectric absorption (A), Raleigh scatter (R), Compton scatter (C), and pair production (P).

If the material composition is exactly known, then the cross sections for the different photon interactions can be calculated using some database, such as XCOM (Berger and Hubbell, 1987) or EPDL97 (Cullen et al., 1997). These databases are available online and they are updated from time to time to include the most recent measurement and calculation results. Most of the general-purpose and specific MC codes are based on one or more of these databases.

Especially in radiation therapy, the material compositions are not always available. Very often, only a CT number (or Hounsfield unit, HU) is known. Therefore, the International Commission on Radiation Units and Measurements (ICRU) has compiled a list of about 100 human tissue types, including material compositions and interaction data (e.g., photon cross sections). These data are published in ICRU Report 46 (Photon, Electron, Proton and Neutron Interaction Data for Body Tissues, 1992). With an adequate CT calibration, an HU number can be mapped to a specific body tissue type. The use of a very limited number of materials, for example, six materials such as air, lung, water, soft tissue, soft bone, and hard bone, should be avoided. Linear interpolation between these supporting points can lead to large cross-section inaccuracies.

The interaction probabilities \(\mu_k(E, \rho)\) \((k = A, R, C, P)\) of Equation 2.23 for human tissue can also be determined more directly, that is, without knowing the material compositions, using an approach presented for MC dose calculation (Kawrakow et al., 1996; Fippel, 1999):

\[
\mu_k(E, \rho) = \frac{\rho}{\rho^*} f_k(\rho) \mu_k^*(E).
\]  

(2.37)

The \(\mu_k^*(E)\) are the interaction coefficients of water and \(\rho^*\) is the mass density of water. As an example, in Figure 2.7, the Compton cross-section ratios \(f_C(\rho)\) for all materials of ICRU Report 46 are plotted. It shows that most of the data points (crosses) are located close to the solid line representing the fit function:

\[
f_C(\rho) = \begin{cases} 
0.99 + 0.01\rho / \rho^*, & \rho \leq \rho^* \\
0.85 + 0.15\rho^*/\rho, & \rho \geq \rho^*.
\end{cases}
\]  

(2.38)
The mixed CH technique as in Section 2.2.2.2 requires knowledge of the hard inelastic collision and bremsstrahlung cross sections $\sigma_{\text{col}}(\vec{r}, E, E')$ and $\sigma_{\text{rad}}(\vec{r}, E, k')$ for energy transfer values $E' > E_c$ and $k' > k_c$, respectively. With Equations 2.28 and 2.40, this allows calculation of the restricted stopping powers according to

$$L_{\text{col}}(\vec{r}, E, E_c) = S_{\text{col}}(\vec{r}, E) - N(\vec{r}) \int_{E_c}^{E} \text{d}E' E' \sigma_{\text{col}}(\vec{r}, E, E')$$

$$L_{\text{rad}}(\vec{r}, E, k_c) = S_{\text{rad}}(\vec{r}, E) - N(\vec{r}) \int_{k_c}^{\infty} \text{d}k' k' \sigma_{\text{rad}}(\vec{r}, E, k').$$

Stopping power tables suitable for MC simulations in radiation therapy have been published, for example, by ICRU for electrons and positrons in Report 37 (Stopping Powers for Electrons and Positrons, 1984), for electrons and protons in Report 46 (Photon, Electron, Proton and Neutron Interaction Data for Body Tissues, 1992), as well as for protons and alpha particles in Report 49 (Stopping Powers and Ranges for Protons and Alpha Particles, 1993). Online databases like ESTAR, PSTAR, and ASTAR (Berger, 1993) can also be used to determine stopping powers for different elements, compounds, and mixtures.

Comparable to Equation 2.37, the stopping powers can also be determined more directly using the mass stopping power ratios $f_s(E, \rho)$ and the mass stopping power of water $S^w_s(E)/\rho^w$ ($k = \text{tot}, \text{col}, \text{rad}$...):

$$S_s(E, \rho) = \frac{D}{\rho^w} f_s(E, \rho) S^w_s(E).$$  

This approach has been adopted, for example, to develop an MC dose calculation algorithm for proton therapy (Fippel and Soukup, 2004). Figure 2.8 shows the total proton stopping power ratio $f_s(E, \rho)$ for two proton energies (10 and 100 MeV) and for all materials of ICRU Report 46 (Photon, Electron, Proton and Neutron Interaction Data for Body Tissues, 1992). After analyzing the tabulations and calculations of ICRU Report 46 (Photon, Electron, Proton and Neutron Interaction Data for Body Tissues, 1992), ICRU Report 49 (Stopping Powers and Ranges for Protons and Alpha Particles, 1993), and the online database PSTAR (Berger, 1993) the following fit formula has been published (Fippel and Soukup, 2004):

$$f_s(E, \rho) = \begin{cases} 
1.0123 & \text{for } \rho \geq 0.9 \\
3.386 \cdot 10^{-5} E + 0.291(1 + E^{-0.3421})(\rho^{0.7} - 1) & \text{for } \rho \leq 0.9
\end{cases}$$

$$f_s(E, \rho) = \begin{cases} 
0.9925 & \text{for } \rho = 0.26 \text{ (lung)} \\
0.8815 & \text{for } \rho = 0.0012 \text{ (air)} \\
\text{interpolate for all other } \rho \leq 0.9
\end{cases}$$

with the kinetic proton energy $E$ in MeV (rest mass not included) and mass density $\rho$ in g cm$^{-3}$. Figure 2.8 shows this function.

The relative uncertainty of function (2.38) is better than 1% for almost all materials of ICRU Report 46. The only exception is gallstone with an error of 1.6%. Similar fits exist for the other interaction types. Thus, it is sufficient to determine only the mass density in a given region and to use equations like (2.38) to calculate the corresponding cross sections.

### 2.2.3.2 Charged Particle Stopping and Scattering Powers

For the MC simulation of charged particle transport using the CH technique according to Section 2.2.2.2, knowledge of the linear restricted collision and radiation stopping powers $L_{\text{col}}(\vec{r}, E, E_c)$ and $L_{\text{rad}}(\vec{r}, E, k_c)$ is necessary. These quantities can be calculated if the unrestricted collision and radiation stopping powers $S_{\text{col}}(\vec{r}, E)$ and $S_{\text{rad}}(\vec{r}, E)$ for the medium at point $\vec{r}$ are known. They provide the average energy loss $\text{d}E$ per step $\text{d}s$ without restricting the energy loss by some threshold. The total unrestricted linear stopping power is defined by

$$S(\vec{r}, E) = -\frac{\text{d}E}{\text{d}s} = S_{\text{col}}(\vec{r}, E) + S_{\text{rad}}(\vec{r}, E).$$

(2.39)

The unrestricted linear collision and radiation stopping powers are calculated using the collision cross section $\sigma_{\text{col}}(\vec{r}, E, E')$ and the bremsstrahlung production cross section $\sigma_{\text{rad}}(\vec{r}, E, k')$ by

$$S_{\text{col}}(\vec{r}, E) = N(\vec{r}) \int_{E}^{\infty} \text{d}E' E' \sigma_{\text{col}}(\vec{r}, E, E')$$

$$S_{\text{rad}}(\vec{r}, E) = N(\vec{r}) \int_{k}^{\infty} \text{d}k' k' \sigma_{\text{rad}}(\vec{r}, E, k').$$

(2.40)

$N(\vec{r})$ is the number of scattering targets per unit volume at point $\vec{r}$ and can be determined by Equation 2.35.

![FIGURE 2.7 Ratios of mass Compton interaction coefficients of all materials from ICRU report 46 to the Compton interaction coefficient of water (crosses). The line represents the fit according to Equation 2.38.](image)
for the two proton energies as solid and dashed lines. Besides some outliers (gallstone, protein, carbohydrate, and urinary stone), Equation 2.43 provides an accuracy of better than 1%, that is, knowledge of the atomic composition in each voxel is not necessary. Equation 2.43 and Figure 2.8 also show an energy dependence of the stopping power ratio can be taken into account by the fit formulas.

Similar formulas exist for electron and positron collision and radiation stopping powers. The material and mass density dependence of the charged particle multiple scattering distributions can also be modeled this way.

References


3.1 Introduction

Monte Carlo (MC) calculations can be time consuming, especially for applications in radiotherapy (RT). Therefore, algorithmic techniques to speed up the simulations are essential. These techniques are called variance reduction techniques (VRT). In this chapter, VRTs with a special focus on RT are introduced and explained. For clarity, it is written from the point of view of photon–electron interactions. However, the same concepts can be applied for heavy charged and neutral particles (protons, neutrons, etc.).

Additional information about VRTs is available, for example, in book chapters by Bielajew and Rogers (1988) or Sheikh-Bagheri et al. (2006), in an article by Kawrakow and Fippel (2000a), and in two task group reports (Chetty et al., 2007; Reynaert et al., 2007). This literature also contains references of the original publications for most of the VRTs presented in the following sections.

3.1.1 Calculation Efficiency

Depending on the number of histories \( N \), the accuracy of any MC calculated mean value \( \langle f(N) \rangle \) of quantity \( f \) is limited by its statistical uncertainty. This uncertainty is given by the variance \( \sigma(N) \) and provides a measure of the statistical fluctuations of the calculated mean value \( \langle f(N) \rangle \) around the true value \( f \) of that quantity. It is obvious that \( \sigma(N) \) decreases with increasing number of histories \( N \) and it becomes zero if \( N \) approaches infinity. In general, \( \sigma(N) \) cannot be calculated because the true value \( f \) is unknown. On the other hand, an estimated variance \( s(N) \) can be calculated during an MC simulation by

\[
s(N) = \sqrt{\frac{\langle f^2(N) \rangle - \langle f(N) \rangle^2}{N - 1}},
\]

with \( \langle f^2(N) \rangle \) being the MC calculated mean of \( f^2 \). The best estimate of the variance is obtained if \( \langle f(N) \rangle \) and \( \langle f^2(N) \rangle \) are calculated using the history-by-history method, that is, they are calculated by averaging over all histories (Salvat et al., 2011). \( \langle f(N) \rangle \) and \( \langle f^2(N) \rangle \) tend to become constant for large numbers of \( N \). Therefore, Equation 3.1 provides a simple method to reduce the variance just by increasing the number of histories \( N \), that is, increasing the calculation time \( T(N) \). However, this is not considered to be a VRT. The purpose of variance reduction is to decrease the time of MC simulations by modifying the algorithm while maintaining an unbiased estimate of the variance \( s(N) \). Unbiased means that for any realistic history number \( N \), the result of MC, including VRT, must not deviate systematically from the corresponding result without VRT.

Instead of VRT, the methods pointed out in this book should be called efficiency enhancement techniques because they improve the efficiency of MC simulations. The calculation efficiency \( \varepsilon \) is defined by


\[ \varepsilon = \frac{1}{[s(N)]^2 T(N)}. \]  

(3.2)

From Equation 3.1, it follows that \([s(N)]^2\) becomes proportional to \(1/N\) for large \(N\). \(T(N)\) is proportional to \(N\). Therefore, the efficiency \(\varepsilon\) is almost independent of \(N\). The calculation efficiency can be improved by reducing the variance \(s(N)\) for a given number of histories \(N\), by decreasing the calculation time \(T(N)\) for a given number of histories \(N\), or by doing both. For historical reasons, we will continue to call these techniques VRT throughout this chapter.

### 3.1.2 Hardware Performance Improvements

The calculation time can be decreased simply by using faster computers or by implementing parallel calculation processes on multicore workstations and computing clusters. However, these methods are not called VRT because they do not make the underlying MC algorithm faster. They just use a given software on a hardware with better performance. Especially, the parallelization of MC calculations is straightforward. Therefore, it is expected that any serious MC algorithm will fully exploit the advantages of present-day computing hardware.

### 3.1.3 Approximate Methods

In the majority of cases, the calculation time per history is decreased by making approximations. Even if the final result is not affected in a significant way, in fact, the approximate methods do not belong to VRTs. According to its original definition, VRTs must not influence the expected result of an infinitely long MC simulation. However, in literature, approximate methods are often denoted as VRT. Some of them, for example, the condensed history technique (CHT), form the basis of almost all MC calculations in radiation therapy today. Therefore, in this chapter, the approximate methods will also be discussed. It is possible to distinguish between real variance reduction techniques (RVRT) and approximate variance reduction techniques (AVRT).

It is obvious that some techniques are AVRTs, for example, the continuous slowing down approximation (CSDA) of the electron transport. On the other hand, it can be difficult to decide whether a given technique belongs to the RVRTs or to the AVRTs. Sometimes, this decision is purely based on intuition or it is shown numerically that some technique is an RVRT. However, numerical experiments are debatable with this respect because the results of both, RVRT and AVRT, must not be influenced significantly. Only by mathematical proof, it can be shown that a definite technique is an RVRT. These proofs can be complex, that is, they would cover a significant amount of space in this chapter. Hence, they will not be shown here. This book focuses on the use of MC techniques in radiation therapy. For both types of VRTs, this means that they have to speed up the simulations without significant loss of numerical accuracy. Therefore, it makes sense in the following sections not to demonstrate for all cases that a specific VRT is an AVRT or an RVRT.

### 3.1.4 Condensed History Electron Transport

The majority of MC algorithms in RT perform electron transport using the CHT. The CHT is an AVRT, and it includes approximations and speeds up MC electron transport significantly compared to analog simulation. Because of its fundamental nature, an in-depth explanation of the CHT is provided in Chapter 2.

### 3.2 Basic Variance Reduction Techniques

In this section, the basic VRTs are outlined. They provide an introduction into the elemental methods of how MC simulation times can be reduced efficiently and what has to be considered to avoid bias of the results. Some of the basic VRTs can be combined to form a more advanced VRT (see Section 3.3).

#### 3.2.1 Uniform Particle Splitting

A common example to explain variance reduction is the particle splitting technique. It can be applied to photons as well as to charged particles. Figure 3.1 demonstrates the splitting technique applied to bremsstrahlung photons. The left drawing represents the MC simulation of an electron hitting the target of a medical linear accelerator (LINAC) and producing one bremsstrahlung photon. For simplicity, we consider the case of only one bremsstrahlung photon because, in general, several photons can be produced by one electron in the target. In a “normal” (also called analog) MC simulation, the photon carries a statistical weight of \(w = 1\). This means that one realistic photon is represented by one photon in the simulation. The right drawing represents the case with a particle splitting factor of \(N_{\text{split}} = 5\), that is, instead of one, five independent bremsstrahlung photons are sampled from the same bremsstrahlung production distribution of the given electron. The VRT is called

![FIGURE 3.1 Schematic representation of an MC simulation of an electron producing one bremsstrahlung photon in the target of a medical linear accelerator (left). On the right-hand side, five bremsstrahlung photons instead of one are created using the splitting technique, each with a statistical weight of \(w = 1/5\).](image-url)
uniform bremsstrahlung splitting (UBS) if all photons are created with the same probability independent of their energy and direction. Then, each of the five photons has to carry a statistical weight of \( w = 1/N_{\text{split}} = 0.2 \), that is, one realistic photon is represented by five photons in the simulation to preserve the total weight.

The reason for bremsstrahlung splitting is given by the purpose of the MC simulation. Two examples are (i) the bremsstrahlung spectrum of the LINAC target shall be calculated or (ii) the target component is part of a full LINAC head simulation with the final goal of dose calculation in a phantom or patient. In both cases, we are interested in a large number of photons to increase the statistical accuracy by lowering the variance. The splitting technique reduces the time required for the creation of five photons in the example of Figure 3.1 because the transport simulation of four additional electrons is saved. In other words, during an analog MC simulation, a lot of time would be “wasted” for the electron transport. However, almost all electrons are absorbed within the target, that is, besides the creation of bremsstrahlung, they do not really contribute to the final result.

Particle splitting is known to be an AVRT, although a proof is not shown here. In the limit of infinite history numbers (number of electrons hitting the target), the results of the analog MC and simulations with photon splitting are identical. This fact is not obvious and it can be wrong if not implemented correctly. An incorrect implementation would arise, for example, if the energy of the electron is reduced by the weighted sum of all split photon energies (five energies in Figure 3.1). Such an approach would bias the shape of the energy spectrum (also known as straggling) of the electrons after the first bremsstrahlung event. Since these electrons can produce secondary and more bremsstrahlung photons, all further results will be influenced by this mistake. To avoid this effect, usually the energy of the electron is reduced just by the energy of one photon arbitrarily selected from all splitting photons. Obviously, the energy is not conserved within one history using this method. On the other hand, the energy conservation law is still fulfilled on an average for large history numbers.

It is important to note, that one should not mix analog MC and splitting or splitting with very different splitting numbers thoughtlessly in one simulation. This could create photons of very different statistical weight, that is, “fat” photons with high weight and “meager” photons with low weight. Only the “fat” photons would contribute to the statistical accuracy of the final result and the time for simulating the “meager” photons would be wasted. A smart implementation of any VRT can produce particles of a very different statistical weight in an intermediate state; however, when the particle properties for the final calculation result are analyzed or processed in a further simulation, it is most efficient if the statistical weight of all the particles is similar.

### 3.2.2 Russian Roulette

Russian roulette can be considered as the opposite of particle splitting. Very often, both techniques are used in combination. In a Russian roulette technique, for a definite particle type (photon or charged particle), a survival probability \( p_{\text{survive}} \) with \( p_{\text{survive}} \ll 1 \) is defined. If a particle of this type is created in an MC simulation, a random number \( \xi \) is sampled from a uniform distribution in interval \([0,1] \). The particle survives if \( \xi < p_{\text{survive}} \), otherwise it is killed, that is, the simulation of this particle stops. To stay in correspondence with reality, the statistical weight of the surviving particles must be increased by the factor \( w = 1/p_{\text{survive}} \).

Russian roulette can be applied efficiently, for example, during the simulation of a LINAC head geometry in combination with the splitting technique described in Section 3.2.1. The bremsstrahlung photons with weight \( w = 1/N_{\text{split}} \) created within the target can hit the secondary collimator of the LINAC. One possible interaction is Compton scattering in the high-Z absorbing material leading to the creation of Compton electrons with weight \( w = 1/N_{\text{split}} \). Most of these electrons are absorbed without producing any secondary (bremsstrahlung) radiation and the calculation time to transport these electrons would be wasted if all of them are simulated. Therefore, it makes sense to kill these electrons with a probability of \( 1 - p_{\text{survive}} = 1 - 1/N_{\text{split}} \). The weight of the surviving electrons must be increased by the factor \( 1/p_{\text{survive}} = N_{\text{split}} \), that is, finally they carry a weight of \( w = 1 \) like the original electrons hitting the bremsstrahlung target. Afterward, it can be useful to again apply splitting if the Compton electrons produce bremsstrahlung. However, in this case, the simulation of these photons makes sense only if they reach the region of interest (ROI). Therefore, some more advanced techniques other than uniform splitting should be the method of choice (see, e.g., Section 3.3.2).

### 3.2.3 Range Rejection

The Compton electrons created in the example of Section 3.2.2 can also be omitted using range rejection instead of Russian roulette. For this VRT, the shortest distance from the present charged particle position to the region boundary must be calculated and compared to the maximum range of that particle in the regions material. Therefore, this technique can only be applied to particles with a definite maximum range depending on the energy, that is, to charged particles. If this range is smaller than the shortest distance to the region boundary, then the charged particle can never leave the present region and it is useful to stop the simulation here.

Obviously, range rejection is an AVRT because a possible production of bremsstrahlung photons is neglected with this approach and photons have a finite probability of leaving the present region. It is of course possible to correctly take these photons into account by adding them via sampling from an approximate distribution.

### 3.2.4 Cross-Section Enhancement

For the simulation of an ion chamber or an air cavity in water with the focus on the energy absorbed in air, it is beneficial to
artificially increase the total photon cross section by some factor \(N_{\text{enhance}}\) in a predefined region around the chamber or cavity. Therefore, the number of photon interactions increase by factor \(N_{\text{enhance}}\) in that region causing a correspondingly increased electron fluence. To maintain an unbiased simulation, the weight of all secondary particles produced in these interactions has to be reduced by multiplication with factor \(w = 1/N_{\text{enhance}}\).

### 3.2.5 Interaction Forcing

Comparable to cross-section enhancement and also applicable to photons only is the interaction-forcing method, schematically represented in Figure 3.2. This is a method that can be used for a photon MC dose calculation engine in RT treatment planning. In Figure 3.2, an incoming photon hits the calculation grid surface at the point denoted as \text{Start}. Then, it is traced along the line denoted with \(s\) until the interaction point. In a worst-case scenario, this point can be outside the calculation grid, that is, behind the \text{Stop} position and the photon does not contribute to the tally. Thus, the time spent on sourcing the photon and transporting it through the calculation grid is wasted. To avoid wasting of calculation time, the photon can be forced to interact between \text{Start} and \text{Stop}.

Interaction forcing is possible if the number of mean free photon path lengths \(\Lambda = \sum_{\text{Start}}^{\text{Stop}} \mu s_i\) between \text{Start} and \text{Stop} can be calculated easily and fast for all crossed voxels \(i\). Here, \(\mu_i\) is the linear attenuation coefficient in voxel \(i\) and \(s_i\) is the photon step length in voxel \(i\). Then, the number of mean free photon path lengths \(\lambda\) can be selected from the distribution function

\[
p(\lambda) d\lambda = \frac{1}{1 - e^{-\lambda}} e^{-\lambda} d\lambda, \quad (3.3)
\]

with \(\lambda\) restricted to the interval \([0, \Lambda]\). That is, with a uniformly distributed random number \(\xi\) from interval \([0,1]\), the number of mean free photon path length \(\lambda\) to the interaction point is calculated by

\[
\lambda = -\ln[1 - \xi(1 - e^{-\Lambda})]. \quad (3.4)
\]

The distance to the interaction point \(P\) is then determined by tracing the photon along the line \(s\) until \(\lambda = \sum_{\text{Start}}^{P} \mu s_i\). Photon forcing requires a weight change of the photon using the factor

\[
w = 1 - e^{-\Lambda}, \quad (3.5)
\]

that is, the photon weight decreases.

In general, each voxel contains a different medium with a different attenuation coefficient \(\mu_i\). Therefore, interaction forcing is useful only if \(\Lambda\) can be calculated fast enough. Otherwise, the MC calculation efficiency can decrease rather than increase.

### 3.2.6 Exponential Transform

The exponential transform works in a similar manner as interaction forcing. Here, the exponential depth distribution of photons is stretched or shortened by some factor \(F > 1\), that is, the number of mean free photon path lengths \(\lambda\) is sampled from the distribution

\[
p(\lambda) d\lambda = \frac{1}{F} e^{-\lambda/F} d\lambda. \quad (3.6)
\]

A uniformly distributed random number \(\xi\) from interval \([0,1]\) provides the number of mean free photon path length \(\lambda\) to the interaction point by

\[
\lambda = -\ln(1 - \xi), \quad (3.7)
\]

that is, for \(F < 1\), \(\lambda\) is shortened, and for \(F > 1\), it is stretched. Exponential transform requires a weight change of the photon using the factor

\[
w = Fe^{-\lambda(F-1)/F} = F(1 - \xi)^{(F-1)/F}, \quad (3.8)
\]

that is, the photon weight is changed depending on the result \(\lambda\) or the random number \(\xi\). This VRT is useful to either increase the calculation efficiency close to the surface (photon entrance point) at the cost of a decreased efficiency in larger depth \((F < 1)\) or vice versa \((F > 1)\).

### 3.2.7 Woodcock Tracking

The basic idea of Woodcock tracking (also called fictitious interaction method) is to make an inhomogeneous simulation

**FIGURE 3.2** Forcing the interaction of one simulated photon within the region of interest, that is, between the \text{Start} and \text{Stop} positions. The number of mean free photon path lengths \(\Lambda = \sum_{\text{Start}}^{\text{Stop}} \mu s_i\) between \text{Start} and \text{Stop} must be calculated for all crossed voxels \(i\) in advance. This can be time consuming because, in general, each voxel contains a different medium with a different attenuation coefficient \(\mu_i\).
geometry homogeneous by adding a fictitious interaction cross section to the total cross section in each region. The fictitious cross sections are chosen in a way to achieve a constant cross section sum for the whole geometry.

For example, to track photons of energy \( E \) in an inhomogeneous material grid, we have to determine the maximum total cross section \( \mu_{\text{max}}(E) \) for the whole simulation geometry. Then, in each region or each voxel with index \( i \), a fictitious interaction cross section is determined by

\[
\mu_{\text{fict}}^{i}(E) = \mu_{\text{max}}(E) - \mu_{\text{tot}}^{i}(E). \tag{3.9}
\]

This allows tracking of photons without ray tracing because of the cross section \( \mu_{\text{max}}(E) \) being independent of the position within the geometry. As soon as the photon has reached the interaction site, a random number can be used to determine the interaction type, real or fictitious. The probabilities for real and fictitious interactions in each voxel are given by

\[
P_{\text{real}}^{i}(E) = \frac{\mu_{\text{real}}^{i}(E)}{\mu_{\text{max}}(E)}. \tag{3.10}
\]

\[
P_{\text{fikt}}^{i}(E) = \frac{\mu_{\text{fict}}^{i}(E)}{\mu_{\text{max}}(E)}. \tag{3.11}
\]

If a real interaction is sampled, the simulation is continued with determining the photon interaction type, for example, Compton scattering or pair production. That is, there is no difference to the conventional MC tracking. However, if a fictitious interaction is sampled, the original photon tracking is continued starting from the present position without changing the photon energy and direction. That is, the fictitious interaction has no effect.

Woodcock tracking implemented in this way does not require weight adjustments of primary and secondary particles. However, it is possible to modify this VRT by introducing weight changes depending on the purpose of the simulation. For example, instead of sampling both real and fictitious interactions, only real interactions are sampled with probability \( P = 1 \). In this case, the weights of all secondary particles produced during these interactions have to be reduced by the factor

\[
w = P_{\text{real}}^{i}(E) = \frac{\mu_{\text{real}}^{i}(E)}{\mu_{\text{max}}(E)}. \tag{3.12}
\]

In combination with Russian roulette, further weight modifications are possible to ensure that the statistical weight of all final particles is comparable.

3.2.8 Correlated Sampling

Applied to the problem of dose calculation in a heterogeneous geometry, correlated sampling starts with a known dose distribution \( D_{\text{hom}}(\vec{r}) \) in a homogeneous phantom, for example, in water. This can be a measured dose distribution, but it can also be a smooth MC dose distribution calculated with very high statistical accuracy. This dose distribution can, for example, be stored in the computer memory. The algorithm then performs a simultaneous calculation of two correlated MC dose distributions, one distribution \( D_{\text{hom}}(\vec{r}) \) in the homogeneous water phantom and another distribution \( D_{\text{het}}(\vec{r}) \) in the heterogeneous geometry. Correlated means that the two simulations are performed using the same sequence of random numbers. To make this calculation fast, the two simulations are performed with low statistical accuracy, that is, with a small number of histories.

The simultaneous simulations are used to determine the distribution of correction factors in each voxel, given by the dose ratios

\[
C(\vec{r}) = \frac{D_{\text{hom}}(\vec{r})}{D_{\text{het}}(\vec{r})}. \tag{3.13}
\]

The final dose distribution in the heterogeneous geometry is then calculated by

\[
D_{\text{het}}(\vec{r}) = C(\vec{r})D_{\text{het}}(\vec{r}). \tag{3.14}
\]

It has been shown that this technique is useful only if the heterogeneous geometry is not too different from the corresponding homogeneous geometry because strong inhomogeneities destroy the correlation between the two simulations.

This technique can also be generalized to two heterogeneous geometries if an accurate solution for one of the geometries is known. An example is the simultaneous calculation of dose to an air cavity volume in water with and without the wall material present.

3.2.9 Initial Calculation of the Primary Interaction Density

For dose calculations with a simple model of the LINAC head, for example, a point source, the density of primary photon interactions can be precalculated in each voxel by performing an initial photon ray tracing. This is schematically shown in Figure 3.3. The MC simulation then starts after this precalculation step with simulating the corresponding number of secondary photons and charged particles in each voxel. The variance is reduced because the primary photon interaction sites are optimally distributed. Investigations have shown that the efficiency can be improved by a factor of about 2 using this method.

The main disadvantage of this method is that it limits the applicability of the code to relatively simple sources. Advanced MC dose calculation engines in RT, however, model the LINAC head using full simulations, using extended virtual source models, or a combination of both. Therefore, an initial calculation of the primary interaction density is not possible for this type of MC applications.
3.2.10 Quasi-Random Numbers

The solution of the transport problem in radiation therapy physics corresponds to an integration in a parameter space with an infinite number of dimensions. If this is performed numerically using the MC method, the dimensionality of the problem has to be limited to a finite number of dimensions, for example, \( n \) dimensions. Using a random number generator, it is possible to sample random vectors in this \( n \)-dimensional space by generating a sequence of \( n \) numbers, forming in this way an \( n \)-tuple. In general, one \( n \)-tuple corresponds to one particle history of the MC simulation. The simulation of \( N \) particle histories corresponds to the generation of a sequence of \( N n \)-tuples. If a normal pseudorandom number generator of good quality is used for that purpose, the distribution of the \( n \)-tuples is suboptimal, that is, the statistical error of the result decreases as \( 1/N \).

Solutions of \( n \)-tuples that fill the \( n \)-dimensional space more uniformly than uncorrelated pseudorandom numbers are called quasi-random sequences. These sequences are known to substantially improve the integration efficiency in certain types of problems. Using quasi-random numbers, the statistical error of the result can decrease asymptotically as \( 1/N \) if the quasi-random \( n \)-tuples are optimally distributed. There are different types of quasi-random numbers; examples are Halton’s sequence or the Sobol sequence (Press et al., 1992).

Similar to pseudorandom numbers, quasi-random numbers are not really random. However, in contrast to pseudorandom numbers, quasi-random numbers cannot be considered as uncorrelated. For example, the Sobol sequence in one dimension starts with 0.5, 0.75, 0.25, 0.375, …, that is, it tries to fill the interval \([0,1]\) in an optimum way. The method is obvious for the one-dimensional case, but it becomes complex for higher dimensions (Press et al. 1992).

3.3 Advanced Variance Reduction Techniques

Advanced VRTs are formed in general by combining different basic VRTs. Some of them have demonstrated to be very efficient for MC calculations in RT, for example, for dose calculations. They are outlined shortly in this section.

3.3.1 Selective Bremsstrahlung Splitting

With UBS (Section 3.2.1), the splitting factor \( N_{\text{split}} \) is a constant throughout the whole simulation, that is, it is independent of the direction of the electron producing the bremsstrahlung photons. With selective bremsstrahlung splitting (SBS) (Sheikh-Bagheri et al., 2006), \( N_{\text{split}}(\theta) \) becomes a function of the variable \( \theta \), the angle between the present direction of the electron and the central beam axis. Because bremsstrahlung photons are forward peaked into a direction that differs only slightly from the direction of the initial electron, it makes sense to use smaller splitting factors \( N_{\text{split}}(\theta) \) with increasing \( \theta \). The maximum should be at \( \theta = 0 \) and the minimum at \( \theta = 180^\circ \). Most of the photons transported with large angles are just absorbed within the primary or secondary collimators of the LINAC head and their calculation time would be wasted. SBS reduces this time waste. Photons with a better chance of reaching the ROI are created with a higher probability using SBS.

There is, however, a serious drawback of this method; it introduces a nonuniform distribution of statistical weights

\[
 w(\theta) = \frac{1}{N_{\text{split}}(\theta)},
\]

whereas \( \theta \) is not the angle of the photon; it is the angle of the original electron. This means that photons in a specific small scoring location can have different weights. This influences the variance in this scoring location and the final efficiency gain of the method is much smaller than expected. SBS improves the efficiency by a factor of 2–3 compared to UBS (Kawrakow et al., 2004).

3.3.2 Directional Bremsstrahlung Splitting

The disadvantages of SBS are eliminated using the directional bremsstrahlung splitting (DBS) method. Unlike UBS and SBS, DBS is a complex algorithm. Hence, the original and systematic publication of Kawrakow et al. (2004) is referred if the reader is interested in details of the method. Here, only a summary will be provided.
DBS uses (as UBS) a constant user-defined splitting factor $N_{\text{split}}$ for the production of bremsstrahlung. Then, the algorithm analyzes the direction of the photons produced. They are transported always, if they are directed into the ROI. If not, Russian roulette (Section 3.2.2) with the survival probability $P_{\text{survive}} = 1/N_{\text{split}}$ is played. Therefore, photons directed into the ROI carry a weight of $w = 1/N_{\text{split}}$. Photons aiming away from the ROI carry a weight of $w = 1$. Additional calculation time can be saved if the directional dependence of the photons is calculated in advance by a smart modification of the bremsstrahlung production cross section (Kawrakow et al., 2004).

By further combining splitting and Russian roulette, by processing “fat” and “meager” particles differently throughout the rest of the simulation as well as by treating photons in gas and higher-density materials differently, it can be ensured that all photons inside the ROI will be “meager,” that is, have a weight of $w = 1/N_{\text{split}}$ and those outside the ROI will be “fat” with a weight of $w = 1$.

A drawback of this default DBS implementation is that all electrons are “fat.” When they contribute to the result, for example, for dose calculation, the electrons would influence the variance in a negative manner. For these purposes, a DBS with electron splitting is available. This technique also ensures that the electrons reaching the ROI are “meager”; however, the overall calculation efficiency decreases compared to the default DBS.

### 3.3.3 Macro Monte Carlo

Macro Monte Carlo (MMC) is based on the premise that the simulation time can be decreased by the use of precomputed results. One of these precomputing techniques for the purpose of dose calculation in RT treatment planning of electron beams is called MMC (Neuenschwander and Born, 1992). As shown in Figure 3.4, this technique is based on precalculated fluence distributions on the surface of spheres for electrons hitting the spheres. A general-purpose MC algorithm can be used to calculate the distributions for incoming electrons of different energy as well as spheres with different diameters and consisting of different materials. The left drawing in Figure 3.4 shows schematically a “full” MC simulation of an electron path within such a homogeneous sphere. With these predictions, the distributions of exit points, angles, and energies are calculated and stored as a function of sphere radius, material, and initial electron energy.

Dose calculation using the patient geometry starts with analyzing the computed tomography (CT) information and processing the calculation grid. Each voxel is assigned with a definite material and a definite sphere diameter, corresponding with the distance till the material changes. In homogeneous regions, spheres with larger diameters can be used. In regions with a strong density fluctuation, smaller diameters must be assigned to the voxels. When an electron hits the calculation grid, the sphere radius and material are determined based on the preprocessing result. The parameters of the electron leaving the sphere are randomly selected from the precomputed macroscopic fluence distributions, that is, the electron directly jumps from the entry to the exit point of the sphere (see the right drawing of Figure 3.4). The energy loss within the sphere is used to calculate the absorbed dose in each voxel.

The process continues with determining the new sphere parameters for the electron leaving the old sphere. It is repeated until the whole electron history is simulated and the energy of the initial electron is consumed.

Compared to conventional MC, MMC is faster by about one order of magnitude. One disadvantage is, however, only the electron transport speed can be increased. This means that MMC cannot be applied to photon beams efficiently.

### 3.3.4 History Repetition

Another technique with recycling of simulation results is called history repetition. This VRT, schematically represented in Figure 3.5, has been developed originally to increase the speed of MC dose calculations for electron beams in RT. Electron history repetition is based on complete electron histories precalculated in a homogeneous water phantom of infinite size. An example of such a precalculated history is shown in Figure 3.5 left from the box. Precalculated means that all parameters of this history (condensed history step lengths, multiple scattering angles, energy losses, secondary particle parameters) are saved for reuse.

On the right-hand side of Figure 3.5, the history is applied to the actual patient geometry $N_{\text{repeat}} = 2$ times, starting at different positions and with different directions at the surface of the calculation matrix. In reality, the optimal $N_{\text{repeat}}$ depends on the size of the electron field. The optimum efficiency gain (approximately a factor of 2) is achieved with about one repetition per 5 cm$^2$ of the field, that is, for a $10 \times 10$ cm$^2$ electron field, the repetition number should be about $N_{\text{repeat}} = 20$. If $N_{\text{repeat}}$ is smaller, the potential of history repetition is not fully exploited. If $N_{\text{repeat}}$
This means that each transport parameter function of type $T$ that depends on the medium $M$ and the electron energy $E$ is approximated as a product of a correction function depending only on the medium and another function depending only on the energy. It has been shown that this is possible for the different types of human tissue, including some phantom materials such as water with an accuracy of 1–2%. With this factorization and without further loss of generality, a reference medium can be selected. In most cases, this is water, that is, for $f_0^T(E)$, the corresponding transport parameter in water is selected:

$$f_0^T(E) = S^T(H_2O,E).$$

Thus, we get

$$S^T(M,E) = f_0^T(M)S^T(H_2O,E).$$

If, for example, $S^T$ is the linear collision stopping power $S^{\text{coll}} = dE/dx$, then for a given infinitesimal small energy loss $dE$, the step length $dx$ according to collision loss is scaled by

$$dx(M) = \frac{dE}{f_0^T(M)S^{\text{coll}}(H_2O,E)} = \frac{dx(H_2O)}{f_0^T(M)}.$$  

The factorization of Equation 3.16 is not a necessary condition for electron histories to be scalable. The correction $f_0^T(M)$ can be extended by including a slight energy dependence. However, electron history repetition does not work for arbitrary materials, for example, metals. Therefore, this technique is an AVRT.

### 3.3.5 Simultaneous Transport of Particle Sets

The limitations of history repetition can be avoided using the simultaneous transport of particle sets (STOPS) technique. With STOPS, several particles of the same type (electron, positron, or photon) and with the same energy are transported simultaneously as a set. As with history repetition, the initial positions, directions, and weights are different. In contrast to history repetition, the histories in one set are not transformed into each other by path length and scattering angle scaling. However, a variety of parameters are independent of the material, so they are sampled once for all particles in the set.

STOPS can be applied for photons as well as for charged particles. Here, electrons are used to explain the technique because STOPS has been developed originally in VMC++ (Kawrakow, 2000; Kawrakow and Fippel, 2000b) to replace electron history repetition. Electrons are transported using a class II CHT (see Chapter 2 for more details) with continuous energy loss to model semielastic collisions as well as discrete Møller (Bhabha for positrons) and bremsstrahlung interactions to model hard inelastic events. The CHT allows one to express distances between discrete interactions as energy.
losses. For the material of type \( M \), they are sampled using the total discrete interaction cross section (number of interactions) per unit energy loss:

\[
\Sigma(E,M,E) = \frac{\Sigma(M,E)}{L(M,E)}.
\]  

(3.20)

\( \Sigma(M,E) \) is the total cross section per unit length and \( L(M,E) \) is the restricted stopping power. An advantage of \( \Sigma(E,M,E) \) is that it depends only weakly on the material type \( M \) as well as on the energy \( E \) and a global maximum \( \Sigma_{E}^{\text{max}} \) for all \( M \) and \( E \) can be efficiently used to perform Woodcock tracking (Section 3.2.7). This means that the geometry becomes homogeneous in terms of number of interactions (fictitious or real) and the energy loss \( \Delta E \) can be sampled once for all electrons within the set. Thus, at the end of the step, all electrons have the same energy.

The geometric step lengths between the initial and final points depend on \( \Delta E \) and on the stopping power \( L(M,E) \). They are calculated separately for each electron in the set because \( L(M,E) \) is different for different materials. Furthermore, the multiple scattering properties are material dependent, that is, the multiple scattering angles are also sampled individually for each particle in the set.

At the end of the step, the interaction type (fictitious, Møller, or bremsstrahlung for electrons) must be determined for each particle separately; however, the same random number can be used for all particles. This causes a very high chance of identical interactions, especially for human tissue because of the weak material dependence of the interaction properties. If the interaction type is identical for all particles, the whole set stays alive. If there are particles with a different interaction type, the set is split into subsets and henceforward the new subsets are transported independently.

If a specific (nonfictitious) discrete interaction is sampled, secondary particle parameters such as the \( \delta \)-electron energy for Møller interactions or the photon energy for bremsstrahlung production must be determined using the differential cross section. The differential Møller cross section is material independent and if we assume the same for bremsstrahlung, the secondary particle energies can be sampled once for the entire set. The polar scattering angles are determined uniquely by the kinematics of the processes. Azimuthal scattering angles always follow a uniform distribution in the interval \([0,360^\circ]\). This distribution is independent of the material type, that is, the azimuthal scattering angle can also be selected once for the particle set. Because identical energies are sampled for equivalent secondary particles within the set, the simulation continues with determining the energy loss to the next interaction by again using the Woodcock scheme. All steps of the simulation are repeated until the entire energy of the set is absorbed or the particles have left the simulation geometry.

The efficiency gain of STOPS is approximately a factor of 2; that is, it is comparable to history repetition. In contrast to history repetition, STOPS can be used for arbitrary media.

3.3.6 Continuous Boundary Crossing

General-purpose MC algorithms usually stop the simulation of charged particles at material interfaces. This is necessary because the underlying physics is valid only for a homogeneous region consisting of one specific material. Especially, the multiple scattering properties are material dependent. Therefore, close to a material boundary, an accurate MC algorithm should switch into a single scattering mode.

However, this is extremely time consuming and for many applications, this is not necessary. For dose calculation in RT treatment planning, history repetition (Section 3.3.4) can be used without loss of accuracy. The multiple scattering angles and step lengths are precalculated in water and scaled depending on the material properties in each voxel. Therefore, it is possible for charged particles to cross boundaries continuously and to trace them efficiently like photons.

3.3.7 Multiple Photon Transport

In combination with photon splitting and Russian roulette, history repetition (Section 3.3.4) and STOPS (Section 3.3.5) can also be used for photon Monte Carlo. To avoid confusion with the various splitting techniques outlined in Sections 3.2.1, 3.3.1, and 3.3.2, here, we call this technique multiple photon transport (MPT). The method has been introduced in Kawrakow and Fippel (2000a) to speed up MC dose calculation for photon beams in RT treatment planning.

Figure 3.6 schematically represents this VRT. The basic idea comes from interaction forcing (Section 3.2.5); the photon is forced to interact within the calculation grid. However, the
time-consuming precalculation of the number of mean free photon path lengths $\lambda$ between the entry and exit points of the grid is avoided with MPT.

With MPT, instead of one photon, a multiple number of $N_{\text{split}}$ identical photons are simulated simultaneously in one ray tracing through the actual geometry. Each of the photons is carrying a weight of $w = 1/N_{\text{split}}$. Identical means that the photons have the same energy, direction, and entry position at the surface of the calculation grid. Only the free path lengths $\lambda_i$ to the first interaction points are sampled differently for the photons. In units of the mean free photon path length, they are calculated using just one random number $\xi$ uniformly distributed in $[0,1]$ by

$$\lambda_i = -\ln \left(1 - \frac{\xi + i}{N_{\text{split}}} \right), \quad i = 0, \ldots, N_{\text{split}} - 1.$$ (3.21)

This means that $\xi$ is used to generate the random numbers for the path length determination of $N_{\text{split}}$ photons, whereas these numbers are distributed with optimum uniformity in interval $[0,1)$ (see Figure 3.7).

Combined with electron history repetition, the MC simulation is performed by repeating the following steps:

1. If $i = 0$, trace primary photons from the cube surface to $\lambda_0$ or trace scattered photons from the interaction site to $\lambda_0$.
2. If $i = 0$, sample one interaction type (Compton scatter, pair production, photoelectric absorption, etc) for all $N_{\text{split}}$ subphotons of the same ray.
3. If $i = 0$, sample secondary particle parameters and charged particle histories in water:
   a. For photoelectric absorption, sample the photoelectron direction, create a corresponding history in water, and store the parameters.
   b. For pair production, sample the electron’s and positron’s energy and direction, create the two histories in water, and store the parameters.
   c. For Compton scattering, sample the scattered particle’s energy and direction, create the Compton electron history in water, and store the parameters.
4. If $i > 0$, trace photons from $\lambda_{i-1}$ to $\lambda_i$ through the inhomogeneous voxel geometry.
5. Apply the charged particle histories to the patient geometry starting at the $i$-th interaction point.
6. In case of Compton scatter, play Russian roulette with the secondary photon using the survival probability

$$p_{\text{survive}} = 1/N_{\text{split}}$$

and store its parameters if it survives.

7. Increase $i$ by 1, that is, take the next subphoton from the ray and go to step 4.

The procedure for the $N_{\text{split}}$ primary photons stops when the cube boundary is reached. Then, the remaining Compton photons are simulated in the same way by executing the MPT function recursively. The procedure with STOPS instead of history repetition works correspondingly.

MPT is fast because $N_{\text{split}}$ photons are transported with only one single ray tracing through the geometry. Furthermore, many quantities are reused with history repetition or STOPS. MPT also reduces the variance because the photon interaction sites are distributed optimally along the ray. With a splitting number of about $N_{\text{split}} = 40$, the MC dose calculation efficiency is increased by a factor of 5–10 using MPT.

### 3.4 Pure Approximate Variance Reduction Techniques

Some of the AVRTs should not really be considered as VRT if the corresponding technique is dominated by one or more approximations. Here, some of these techniques, called pure AVRTs, are presented.

#### 3.4.1 KERMA Approximation

An approximation based on KERMA is motivated by the fact that the energy released by low-energy photons is absorbed in the direct neighborhood of the photon ray. For these photons, the MC transport of secondary electrons can be switched off. That is, for all voxels along the photon ray, the deposited energy is calculated by

$$\Delta E_{\text{dep}} = E\mu_{\text{en}}^M(E)\Delta s.$$ (3.22)

Here, $\Delta s$ is the length of the photon path within the given voxel, $E$ is the energy of the photon, and $\mu_{\text{en}}^M(E)$ is the linear energy absorption coefficient for photons of energy $E$ and the material $M$ in the corresponding voxel.

Systematic deviations of the dose distributions can be minimized with this approach if the maximum range of the secondary electrons is smaller than the spatial resolution of the calculation grid. This is possible if the KERMA approximation is applied only to photons with an energy $E$ smaller than some predefined maximum energy $K_{\text{cut}}$. Furthermore, it is useful to apply this technique only for secondary and higher-order scattered photons, but not to primary photons coming directly from the LINAC head.

#### 3.4.2 Continuous Slowing Down Approximation

In CSDA, electrons (or other charged particles) are transported in one step without the creation of secondary particles. The length of the step is calculated from the energy of the electron...
and the unrestricted total stopping power. This continuous slowing down range is a unique function of the energy of the electron. Therefore, energy straggling is neglected using this approach.

There are versions of CSDA with and without multiple scattering. Without multiple scattering, the charged particle is transported on a straight line. Otherwise, a multiple scattering angle is sampled and included in the simulation.

CSDA should be used only for low-energy charged particles, for example, for electrons within an MC simulation, when their energy drops below some user-defined threshold energy $E_{\text{cut}}$. However, CSDA for subthreshold electrons is more accurate and produces less statistical fluctuations than local absorption of the track-end energy.

### 3.4.3 Transport Parameter Optimization

Both the accuracy and speed of coupled electron–photon transport MC simulations in RT depend on the selection of various transport parameters such as particle transport cut-off energies, particle production threshold energies, and condensed history step sizes.

MC algorithms usually employ a photon energy cut-off parameter $P_{\text{cut}}$, that is, photons are not transported if they are generated within the simulation with an energy below $P_{\text{cut}}$. The remaining energy can be neglected or it can be deposited locally. It is obvious that the accuracy of the result increases with decreasing $P_{\text{cut}}$, however, the calculation time also increases. On the other hand, if the remaining photon energy is absorbed locally, a large value for $P_{\text{cut}}$ can cause additional fluctuations. This can decrease the calculation efficiency, although the calculation time has also decreased.

Another important parameter is the photon production threshold energy $P_{\text{min}}$, for example, for bremsstrahlung. This means that only bremsstrahlung photons with an energy larger than $P_{\text{min}}$ can be generated during the MC simulation. The effect of electron stopping due to bremsstrahlung photons below $P_{\text{min}}$ is taken into account by the restricted radiative stopping power. The question here is how to deal with the radiative energy released by these electrons? Both local absorption and total disappearance of this energy are approximations. To avoid a significant influence on the result on the one hand and to ensure short calculation times on the other, $P_{\text{min}}$ should be selected with care depending on the type of the calculation.

The problem is less complex for the cut-off and production threshold energies of charged particles, $E_{\text{cut}}$ and $E_{\text{min}}$. Because of the strong correlation between energy and range, both parameters can be selected depending on the spatial resolution of the calculation geometry. On the other hand, efficiency can be improved by using a higher value of $E_{\text{cut}}$ and by simulating track-end electrons with energy below $E_{\text{cut}}$ in CSDA (Section 3.4.2).

Another important parameter in condensed history MC electron algorithms is the step size (or the procedure used to determine the actual step size depending on energy, material, geometry, etc.). The step size must be restricted by some maximum if an approximate multiple scattering theory, for example, a small-angle approximation is implemented. In widespread use are approaches with a user-defined maximum energy loss per step parameter $E_{\text{step}}$. The value of $E_{\text{step}}$ should be optimized depending on the algorithm and the type of the application.

### References


## Application of Monte Carlo Techniques in Radiation Therapy

<table>
<thead>
<tr>
<th>4</th>
<th>Applications of Monte Carlo to Radiation Dosimetry</th>
<th>Hugo Bouchard and Jan Seuntjens</th>
<th>43</th>
</tr>
</thead>
<tbody>
<tr>
<td>5</td>
<td>Monte Carlo Modeling of External Photon Beams in Radiotherapy</td>
<td>Frank Verhaegen</td>
<td>63</td>
</tr>
<tr>
<td>6</td>
<td>Monte Carlo Modeling of External Electron Beams in Radiotherapy</td>
<td>Frank Verhaegen</td>
<td>87</td>
</tr>
<tr>
<td>7</td>
<td>Dynamic Beam Delivery and 4D Monte Carlo</td>
<td>Emily Heath and Joao Seco</td>
<td>95</td>
</tr>
<tr>
<td>8</td>
<td>Patient Dose Calculation</td>
<td>Joao Seco and Maggy Fragoso</td>
<td>111</td>
</tr>
<tr>
<td>9</td>
<td>Monte Carlo Methods and Applications for Brachytherapy Dosimetry and Treatment Planning</td>
<td>Guillaume Landry, Mark J. Rivard, Jeffrey F. Williamson, and Frank Verhaegen</td>
<td>125</td>
</tr>
<tr>
<td>10</td>
<td>Monte Carlo as a QA Tool for Advanced Radiation Therapy</td>
<td>JinSheng Li and C.-M. Charlie Ma</td>
<td>145</td>
</tr>
<tr>
<td>11</td>
<td>Electrons: Clinical Considerations and Applications</td>
<td>Joanna E. Cygler and George X. Ding</td>
<td>155</td>
</tr>
<tr>
<td>12</td>
<td>Photons: Clinical Considerations and Applications</td>
<td>Michael K. Fix</td>
<td>167</td>
</tr>
<tr>
<td>13</td>
<td>Monte Carlo Calculations for Proton and Ion Beam Dosimetry</td>
<td>Hugo Palmans</td>
<td>185</td>
</tr>
<tr>
<td>14</td>
<td>Protons: Clinical Considerations and Applications</td>
<td>Harald Paganetti</td>
<td>201</td>
</tr>
</tbody>
</table>
15 Application of Monte Carlo Methods to Radionuclide Therapy  Michael Ljungberg .......................................................... 223
Introduction • MIRD Formalism • Scintillation Camera • Activity Quantitation • Absorbed Dose Calculations • Diagnostic Dosimetry • Therapeutic Dosimetry • Monte Carlo Codes for Simulating Nuclear Medicine Systems • Voxel-Based Phantoms • Summary • References

16 Monte Carlo for Kilovoltage and Megavoltage Imaging  George X. Ding and Andrew Fielding......................... 233
Introduction • Monte Carlo and Kilovoltage X-Ray Imaging • Monte Carlo and MV Imaging • References

17 Monte Carlo Calculations for PET-Based Treatment Verification of Ion Beam Therapy  Katia Parodi............... 251
Introduction • Monte Carlo Modeling of Ion-Induced β+-Emitters • Monte Carlo Modeling of Imaging Process • Remaining Challenges for Clinical Implementation • Conclusion and Outlook • References

18 Monte Carlo Studies of Prompt Gamma Emission and of Proton Radiography/Proton-CT  Jerimy C. Polf, Nicholas Depauw, and Joao Seco ............................................................... 263
State of the Art in Prompt Gamma Research • Basic Properties of PG Emission • Range Verification and Tissue Composition Analysis • Detector Design Studies • Image Reconstruction Studies • Validation of Monte Carlo-Based PG Calculations • State of the Art of Proton Radiography • Protons versus X-Ray • Proton Radiography Image Reconstruction Differences • Spatial Resolution and Density Resolution • Proton Radiography, kV and MV Image Quality for Identical Integral Dose • Proton Radiography: A Step toward Proton-CT • References

19 Monte Carlo for Treatment Device Design  Bruce A. Faddegon .............................................................. 273
Introduction • Target for Therapy • Target for Imaging • Electron Beam Magnetic Shield • X-Ray Standard Flattener and Extended Flattener • X-Ray Leakage • Electron Secondary Scattering Foils • X-Ray Wedges • Electron Applicator Leakage • Electron Multileaf Collimator • X-Ray Image Detection and Processing • Conclusion • References

20 GPU-Based Fast Monte Carlo Simulation for Radiotherapy Dose Calculation  Xun Jia, Sami Hissoiny, and Steve B. Jiang ................................................................. 283
Introduction • Graphics Processing Unit: Hardware and Programming • Monte Carlo Simulation on GPU for Radiotherapy • Current GPU-Based Monte Carlo Codes • Summary • References

21 Monte Carlo for Shielding of Radiotherapy Facilities  Peter J. Biggs and Stephen F. Kry................................. 295
Introduction • Monte Carlo Codes • Photon Shielding • Neutrons from High-Energy Photon Therapy • Application of Monte Carlo for Nonstandard Room Geometries • Nonstandard Beams and Accelerators • Summary and Future Directions • References
4

Applications of Monte Carlo to Radiation Dosimetry

4.1 Introduction and Scope

Over the last decades, developments in Monte Carlo radiation transport algorithms have had tremendous impact in different areas of radiation dosimetry. During the experimental determination of absorbed dose, several quantities are often difficult (or impossible) to estimate accurately without numerical models. In the context of radiation dosimetry, detectors are most of the time constituted of several components, whose material differs substantially from the medium where absorbed dose is to be known. This situation induces a well-known problem that can be characterized in terms of perturbation factors and changes in response to energy (or energy dependence). Over the different generations of radiation dosimetry protocols, progress in Monte Carlo techniques has affected the improvement in accuracy during the determination of absorbed dose. While air kerma-based protocols of the 1980s, for example AAPM's TG-21 (AAPM, 1983), were the first ones to consider the ionization chamber design in detail, some quantities such as the replacement and the wall perturbation factors (i.e., $P_{\text{rep}}$ and $P_{\text{wall}}$, respectively) were estimated with approximate measurements, analytical models, or simply taken as unity based on judicious assumptions. Nowadays, such factors can be calculated with Monte Carlo methods without the need for such approximations, improving the accuracy of absorbed dose determination with radiation detectors.

This chapter presents an overview of the different applications of Monte Carlo to photon and electron beam radiation dosimetry for the energy range used in radiation therapy. Following the introduction, the basics of measurement dosimetry is defined, including physical quantities and units as well as the formalism of reference absorbed dose determination. In the third section, the calculation of quantities related to the standard formalism of absorbed dose measurement is described. The fourth section summarizes several studies using detector response simulations in a context relevant to clinical dosimetry and describes the tools currently available to perform such studies. The fifth section treats the topic of Monte Carlo simulation accuracy in measurement dosimetry, describing the different limitations contributing to that issue. In the sixth section, the evolution of nonstandard beam dosimetry techniques is treated with emphasis on the role of Monte Carlo in the acquisition of dosimetry-related data. Finally, a conclusion is made describing the expected role of the Monte Carlo method in the future of radiation dosimetry.

4.1 Introduction and Scope ........................................................................................................... 43
4.2 Basics of Measurement Dosimetry of Photon and Electron Beams ................................. 44
   Historical Context of Cavity Theory • Theory of Measurement Dosimetry • In-Phantom Dosimetry • Air Kerma Dosimetry
4.3 Calculation of Radiation Dosimetry-Related Quantities....................................................47
   Stopping-Power Ratios • Interaction Coefficients
4.4 Modeling Detector Response................................................................................................ .. 48
   Ionization Chamber Correction Factors • Simulation Geometries • Efficiency Improvements • Positioning Uncertainties in Detector Response Calculations
4.5 Accuracy of Detector Response Calculations .......................................................................52
   General Considerations • Accurate Condensed History Technique • Fano Test • Accuracy of Physical Data • Effects of Machine Modeling • Size of Volume of Interest • Effects of Geometry Accuracy
4.6 Small and Nonstandard Field Dosimetry..............................................................................55
   Small Fields • Intensity-Modulated Beams
4.7 Conclusions................................................................................................................ .................57
References ...............................................................................................................................................57
4.2 Basics of Measurement Dosimetry of Photon and Electron Beams

4.2.1 Historical Context of Cavity Theory

In 1936, Gray published one of the first papers dealing with the fundamental problem of converting absorbed dose in an air-filled cavity to absorbed dose in water (Gray, 1936), including earlier work by W. L. Bragg (1912) and himself (Gray, 1929). In their approach, an idealistic situation is defined where a volume of interest achieves two artificial conditions implying that the amount of energy lost by electrons through electronic collisions is equal to the energy absorbed in that cavity. As a consequence, the expression of absorbed dose in photon (or electron) beams is reduced to absorbed dose to gas in the cavity times the ratio of stopping powers, which depends on the beam energy (Laurence, 1937). Spencer and Attix (1955) took a closer look at the Bragg–Gray formulation in an attempt to solve discrepancies between experiments and predictions due to a breakdown of the conditions required by the principle. To account for “fast” secondary electrons (delta rays) that could exit the cavity, Spencer and Attix defined a two-group energy deposition schematization where electronic inelastic collisions are considered dissipative only if their kinetic energy is below a threshold value (also known as cut-off energy), denoted Δ. The electron slowing down fluence spectrum was derived by solving the Boltzmann transport equation for infinite media (see, e.g., McGinnies, 1959) and based on that a new formulation of stopping-power ratio was defined, providing a better agreement to experiments than the Bragg–Gray formulation.

With technological developments, Monte Carlo methods emerged as a solution to the radiation transport problem allowing accurate estimation of the electron spectra within media (Berger, 1963). Despite the treatment by Spencer and Attix as well as Burch (1955, 1957), the conditions of the Bragg–Gray theory formed the basis of radiation dosimetry protocols in the 1960s (e.g., the ICRU Report no. 14, 1969) with Spencer–Attix stopping-power ratios, adopted in the 1970s (e.g., the ICRU Report no. 21, 1972). Significant theoretical progress in ionization chamber dosimetry was made by Almond and Svensson (1977), whereas progress in Monte Carlo techniques allowed for practical calculation of stopping-power ratios. Nahum (1978) published a modified formulation of stopping-power ratios suitable for use with Monte Carlo methods. By adding a track-end term to the Spencer–Attix’s formulation, as reported earlier by Burch (1957), stopping-power ratio data were further calculated extensively based on this modified formulation (e.g., Andreo and Nahum, 1985). Despite the fact that other improved approaches to cavity theory were also proposed (e.g., Burlin, 1966; Janssens et al., 1974; Zheng-Ming, 1980; Janssens, 1981; Bielajew, 1986), the radiation dosimetry community widely adopted the Spencer–Attix–Nahum formulation of stopping-power ratios, being applicable to megavoltage photon and electron beams (e.g., the ICRU Report no. 35, 1984). For several decades, the Spencer–Attix–Nahum formulation has been extensively included in dosimetry protocols (e.g., AAPM’s TG-21, 1983; IAEA’s TRS-277, 1987), air kerma standards, electron beam dosimetry (e.g., AAPM’s TG-25, 1991), and beam quality correction factors (e.g., AAPM’s TG-51, 1999; IAEA’s TRS-398, 2001).

4.2.2 Theory of Measurement Dosimetry

The determination of absorbed dose is performed using a radiation detector in a reference dosimetry setup, defined by the radiation source and geometry of interest. By performing an experimental measurement of the detector reading or signal in these conditions, one can obtain the absorbed dose to the medium of interest by converting the detector signal, being induced by a physical process specific to the device, to absorbed dose to the medium at the point of measurement. In the context of this review, we will consider that appropriate corrections have been applied to the detector signal, so as to (1) refer to standard reference conditions (e.g., environmental conditions in ionization chambers) and (2) correct for effects that influence the measurement of the physical process itself (e.g., ionization chamber recombination and polarity effects). We call the result the detector signal. Note that Monte Carlo calculations can also play a role in the determination of these corrections (see, e.g., Abdel-Rahman et al., 2006), for the polarity effect, but we will not review these studies further as part of this review.

The detector signal is first converted to absorbed dose in the sensitive volume of the detector, also called the detector cavity. For a given detector signal \( M_{\text{det}} \), the absorbed dose in the detector cavity, denoted \( D_{\text{det}} \), can be expressed as follows:

\[
D_{\text{det}} = C_D M_{\text{det}}, \tag{4.1}
\]

where \( C_D \) is a global factor representing the conversion of detector signal to absorbed dose in the detector cavity for a specific beam quality \( Q \). The factor depends on the physical detection mechanism and incorporates the intrinsic energy dependence of the detector, its linearity with dose and its dose rate dependence. Note that \( C_D \) can also be expressed as the product of a series of factors, each accounting for these effects (see, e.g., Rogers, 2009). The intrinsic energy dependence describes the link between the physical effect and the dose to the detector sensitive volume. The determination of the intrinsic energy dependence thus requires models of the physical or chemical behavior of the mechanism that links the observed reading to the absorbed dose to the sensitive volume. For example, Klassen et al. (1999) described the energy dependence of the G-value of Fricke dosimeters in megavoltage beams. In practice, the intrinsic energy dependence is usually determined indirectly, by calibration against a standard at a specific beam quality \( Q \) (e.g., \(^{60}\)Co for linac radiotherapy) and after extracting the extrinsic energy dependence as calculated by condensed history (CH) Monte Carlo methods (see below). The importance of differentiating intrinsic and extrinsic energy dependence lies in the ability to determine the impact of beam quality change on the response of a detector arising from fundamental effects (i.e., physical signal to detector dose conversion) versus dosimetric effects,
arising from detector embedding, walls, and so on, that can be modeled through CH Monte Carlo techniques.

Typical point-dose detectors are constituted of a homogeneous detection cavity (the sensitive volume) as well as other structural components being designed to both assure the physical integrity of the detector and ensure basic operation. As the radiation field interacts with these nonsensitive components, the result from these interactions is particular to the detector geometry and materials where such components are constituted differently from the medium in which the detector lies. Therefore, since the measured cavity response differs from what would be measured without these components being present, the assessed quantity is perturbed. This ensemble of phenomena can be termed the component-specific perturbation effects.

For a detection cavity freed of nonsensitive components, there remains some perturbation effect during radiation detection, since the cavity is also (most of the time) constituted of a material different from the medium. As physical properties such as mass density, atomic number, and the mean ionization potential (I-value) affect the charged particle fluence differential in energy and direction, which is a result of interactions in the geometry, the energy fluence in the cavity differs from the one at the point of measurement without the detector being present. This overall phenomenon is defined as the detector replacement perturbation effect. This means that if, hypothetically, one replaces the detection cavity, so far assumed freed of nonsensitive components, with an infinitely small detection cavity (or small enough within the limits of the definition of absorbed dose), the charged particle fluence at the point of measurement would be the same as the one in the medium without any detector in place. Similarly, as the detector perturbation effect, the replacement perturbation effect depends on the cavity size and shape, the nature of the material constituting the cavity, as well as the irradiation conditions.

Note that historically, radiation dosimetry was based on Bragg–Gray principles and Fano’s theorem (Fano, 1954), which implies that there is no replacement perturbation effect under charged particle equilibrium (CPE) conditions. For practical situations, Fano’s theorem is not strictly applicable (Bjårrangard and Kase, 1985; Bouchard et al., 2012) and therefore a replacement perturbation effect is always present, especially for low energies where electron scattering power is high.

While the perturbation effects are also energy dependent, the overall conversion of absorbed dose in the detection cavity to absorbed dose to medium, considering the detector in its entire physical integrity, is written as follows (Rogers, 2009):

\[ D_{\text{med}} = f D_{\text{det}}, \quad (4.2) \]

where \( D_{\text{med}} \) is the absorbed dose to the medium at the point of measurement and \( f \) is a factor representing the overall absorbed-dose energy dependence of the detector (also called the extrinsic energy dependence). This factor encompasses the component-specific and replacement perturbation effects as well as the absorbed-dose energy dependence of the detection material.

Although the component-specific and replacement perturbation effects must be evaluated to accurately determine absorbed dose to the medium in which the detector is placed in, the absorbed-dose energy dependence of the detection material relative to the medium must also be assessed. During radiation transport, the density of ionizations in charged particle tracks is much higher than for photons, which in principle interact only once, and therefore the contribution of charged particles to dose is dominant, whether the primary beam is a photon or electron beam. Furthermore, as electrons are set in motion by ionization, their fluence is higher than for positrons, typically by at least a few orders of magnitude for radiotherapy beams. Therefore, collision mass stopping power averaged over the electron spectrum is a major indicator of the absorbed-dose energy dependence of the detection material, with respect to the overall solution of Boltzmann radiation transport equations achieved by the Monte Carlo method (see, e.g., Bouchard, 2012). This means that even for an infinitely small detection cavity, freed of the effect of components, the conversion of absorbed dose to the detector to absorbed dose to medium must consider the energy dependence of the mean mass stopping power, which is defined by the electron fluence spectrum and the physical properties of the media involved.

The detector response (sometimes termed detector sensitivity) can be defined as the ratio of its reading, \( M_{\text{det}} \), to the quantity of interest, which could be air kerma or dose to a medium. For example, the detector absorbed dose response or sensitivity is defined as

\[ R_D = \frac{M_{\text{det}}}{D_{\text{med}}}, \quad (4.3) \]

where \( M_{\text{det}} \) is the reading of the detector placed in the medium \( D_{\text{med}} \) the absorbed dose at the effective point of measurement (EPOM) in the medium \( \text{med} \). By combining Equations 4.1 and 4.2 into Equation 4.3, the absorbed dose response can be written as

\[ R_D = \frac{1}{f C_Q}, \quad (4.4) \]

As discussed above, Monte Carlo simulations of detector response are involved with the extrinsic component of the response, that is, the calculation of \( f \).

### 4.2.3 In-Phantom Dosimetry

Absorbed dose measurement techniques are specific to a given detector, geometry, and beam quality. In the context of radiation dosimetry, the beam quality should be viewed as a beam-specific and geometry-dependent concept, rather than a descriptive of the beam only. It is defined as the overall condition induced by a given radiation beam in a specific region of the geometry of interest. While this condition is multidimensional, quality can be related to dosimetric factors being detector specific and
position dependent. In practice, beam quality is encapsulated in a single quantity known as the in-water beam quality specifier (e.g., AAPM TG-51’s %dd(10), (Almond et al., 1999)). This quantity is used in specific applications where dosimetric factors need to be characterized as a function of beam quality. The function $C_d$, the absorbed-dose energy dependence factor, and $f$, defined in Equations 4.1 and 4.2, respectively, depend on the nature of the detector and the beam quality. While the standard formalism is valid for any detector, it was historically developed for ionization chambers, the gold standard in clinical reference dosimetry. For a given beam quality, the standard formalism to convert detection signal to absorbed dose in the detection cavity is as follows:

$$C_d = N_{gas},$$

where $N_{gas}$ is a signal-to-dose calibration coefficient that can be expressed as

$$N_{gas} = \frac{(W/e)_{gas}}{m_{gas}}.$$

Here, $(W/e)_{gas}$ is the energy necessary to create an ion pair in the gas filling the chamber for a cobalt-60 beam (e.g., 33.97 eV/e for sea-level dry air (Boutillon and Perroche-Roux, 1987)) and $m_{gas}$ is the mass of gas (air) in the effective sensitive volume of the chamber. The absorbed-dose energy dependence factor $f$, used to convert detector dose to dose to medium, is expressed as

$$f = \left(\frac{L}{\rho}\right)_{gas}^{med} \rho_{repl} P_{wall} P_{cel} P_{stem},$$

where $(L/\rho)_{gas}^{med}$ is the medium-to-gas Spencer-Attix stopping-power ratio, representing the absorbed-dose energy dependence of the detection material, $P_{repl}$ is the replacement perturbation factor, and $P_{wall}$, $P_{cel}$, and $P_{stem}$ are the wall, central electrode, and stem perturbation factors, respectively.

In a more general context, the standard formalism can be generalized to any detector. While the approach to convert detector signal to absorbed dose is linear for ionization chambers, it can be stated differently and specifically to the detection mechanism. As for the absorbed-dose energy dependence factor stated in Equation 4.7, the component-specific perturbation factors (i.e., $P_{wall}$, $P_{cel}$, and $P_{stem}$) are clearly defined by the presence of these components. Generally, there could be a series of perturbation factors that require attention to account for all nonsensitive components constituting the detector. That is, one can define the following for a more general context:

$$f = \left(\frac{L}{\rho}\right)_{det}^{med} P_{repl} \prod_i P_i,$$

where $P_i$ are the component-specific perturbation factors of the detector.

### 4.2.4 Air Kerma Dosimetry

Air kerma measurements with ionization chambers have been, for several decades, fundamental to clinical radiation dosimetry (e.g., AAPM’s TG21, 1983; IAEA’s TRS-277, 1987). The development of calorimetry techniques (Ross and Klassen, 1996; Seuntjens and DuSautoy, 2003; Seuntjens and Duane, 2009) allowed the development of absorbed-dose-to-water-based dosimetry protocols (TG-51, TRS-398). This approach has become, over the last decade, the standard method for megavoltage beam dosimetry. Despite this change of practice, air kerma standards remain essential for the calibration of ionization chambers for kilovoltage photon beam dosimetry (e.g., AAPM’s TG-61 (Ma et al., 1999)).

The rationale behind air kerma measurements is based on Fano’s theorem (Fano, 1954), stating that under the condition of CPE, the charged particle fluence differential in energy and direction is independent of point-to-point density variations in the medium. If one defines a geometry in which an air-filled cavity is irradiated by a photon beam, Fano’s theorem implies that the absorbed dose equals collision kerma (see illustration in Johns and Cunningham’s book, 1983).

In practice, a thick wall of air-equivalent material is chosen to enclose the air cavity, such that the perturbation of the electron spectrum between the wall and the air is minimal. The beam quality-specific wall thickness is chosen larger than the electron range (in the wall material), such that electron equilibrium (strictly speaking, a transient equilibrium) is established between the wall and secondary electrons in the wall. Using Fano’s theorem, one can assume that the electron fluence perturbation by the presence of the cavity is negligible, this statement corresponding to the first condition required for Bragg–Gray dosimetry (which is also required in the Spencer–Attix cavity theory). The second required condition, that is, that photon interactions in the cavity are negligible, can also be assumed. This way, one can state that the absorbed dose to a mass of wall occupying the chamber sensitive volume is expressed as

$$D_{cav,wall} = \left(\frac{L}{\rho}\right)^{wall} D_{cav,air},$$

where $D_{cav,air}$ is the absorbed dose in the detection cavity filled with air.

The next step of the rationale is to convert absorbed dose in the detection cavity filled with wall material into absorbed
dose to the medium consisting the geometry. Assuming that the attenuation of the photon fluence in the wall and the medium are identical, this is achieved by simply taking the ratio of collision kerma wall-to-medium, as follows:

\[
K_{\text{coll,med}}^{\text{no att. or scat.}} = \left( \frac{\mu_{\text{med}}}{\mu_{\text{wall}}} \right)_{\text{wall}} D_{\text{cav,wall}},
\]

(4.10)

where \(\left(\frac{\mu_{\text{med}}}{\rho}\right)_{\text{wall}}\) is the ratio of mass-energy absorption coefficients, medium-to-wall. In Equation 4.10, it is implicitly assumed that the photon energy fluence in medium (air) and wall is identical, and that there is perfect CPE in the wall (i.e., \(K_{\text{coll,wall}} = D_{\text{wall}}\)). In the context of realistic detectors, one admits that (1) CPE is not perfectly achieved and (2) the presence of the wall causes perturbation in photon fluence. In a more general manner, one can state the absorbed-dose energy dependence factor \(f\) as follows:

\[
f = \left( \frac{L}{\rho} \right)_{\text{wall}} \left( \frac{\mu_{\text{med}}}{\rho} \right)_{\text{wall}} K_{\text{wall}},
\]

(4.11)

where \(K_{\text{wall}}\) is a correction factor accounting for differences in photon and electron attenuation and scatter between the wall and the medium, and \(K\) is a general correction factor accounting for the fact that Fano’s theorem is not entirely applicable (i.e., CPE is not established or the wall is not perfectly air-equivalent). The factor \(K\) also accounts for other component-specific effects (e.g., a build-up cap, a central electrode), as well as the fact that the chamber does not determine air kerma at a point, while this is what is meant to be measured, due to the anisotropy of the source and the nonlinear fluence-to-distance relation.

The final step of the method consists of choosing the medium as air, and to convert absorbed dose in free air to air kerma. The result yields the following expression:

\[
K_{\text{air}} = \frac{1}{\left(1 - g_{\text{air}}\right)} f \cdot D_{\text{cav,air}},
\]

(4.12)

where \(1 - g_{\text{air}}\) is the ratio between collision kerma and kerma, accounting for the kinetic energy lost in radiative interactions (bremsstrahlung, annihilation-in-flight).

4.3 Calculation of Radiation Dosimetry-Related Quantities

4.3.1 Stopping-Power Ratios

As described in the above section, Spencer–Attix cavity theory is fundamental to radiation dosimetry of photons and electron beams. The expression of medium-to-detector stopping-power ratio is given by Spencer–Attix–Nahum’s formulation as follows (Spencer and Attix, 1955; Nahum, 1978):

\[
\left( \frac{L}{\rho} \right)_{\text{det}}^{\text{med}} = \int_{\Delta}^{\infty} \varphi_{\text{e}}(T)(I_{\text{med}}(T)/\rho_{\text{med}}) \, dT + T E_{\text{med}}
\]

\[
= \int_{\Delta}^{\infty} \varphi_{\text{e}}(T)(I_{\text{det}}(T)/\rho_{\text{det}}) \, dT + T E_{\text{det}},
\]

(4.13)

where \(\varphi_{\text{e}}(T)\) is the electron fluence differential in energy (in MeV^{-1} cm^{-2}). Here, for either material (i.e., the medium or the detector), \(L(T)\) is the restricted collision stopping power, \(\rho\) is the mass density, det is the medium of the sensitive volume of the detector, med is the medium the detector is placed in, and \(TE\) is the track-end term, discussed shortly. Note here that as described by Spencer–Attix–Nahum cavity theory (which assumes a detector satisfying Bragg–Gray conditions), the electron fluence \(\varphi_{\text{e}}(T)\) is by definition unperturbed by the presence of the detector, and therefore corresponds to the electron fluence in the medium without any detector being present.

There are two ways in which stopping-power ratios are calculated. The first method consists of calculating stopping-power ratios analytically, using Equation 4.13. In such an approach, a Monte Carlo simulation of the transport through the medium of interest is used to estimate the electron fluence \(\varphi_{\text{e}}(T)\), differential in energy, by calculating the average path length in the cavity over its volume as a function of kinetic energy. The track-end term in the analytical method is defined as follows (Nahum, 1978):

\[
T E_{\text{med}} = \varphi_{\text{e}}(\Delta) \left( \frac{S_{\text{coll}}(\Delta)}{\rho_{\text{med}}} \right) \Delta,
\]

(4.14)

where \(S_{\text{coll}}\) is the collision electron stopping power in the medium in question. Obviously for such approach, the choice of the simulation energy threshold, which separates discrete interactions and condensed steps, must be smaller than \(\Delta\).

The second and more modern approach to stopping-power ratio calculations (see SPRRZNR’s manual (Rogers et al., 2011)) is a direct implementation of Equation 4.13 in the code, avoiding an offline calculation and the computation and storage of electron fluence spectra. This is particularly relevant if stopping-power ratios are calculated for several positions in the geometry. The radiation transport is performed in a single medium. During the simulation, the electron condensed step and electron track ends are scored in two ways: (1) in the medium, where the transport is determined from physical properties of the geometry, and (2) in the detection material, where the absorbed dose is scaled by the ratio of restricted or unrestricted stopping powers. During a CH step, the absorbed dose in the detection material is given by

\[
D_{\text{DEP, det}} = D_{\text{DEP, med}} \left( \frac{I_{\text{det}}(T)/\rho_{\text{det}}}{I_{\text{med}}(T)/\rho_{\text{med}}} \right),
\]

(4.15)

where DDEP represents the absorbed dose during an electron step. Here, the numerator and denominator are the restricted
collision stopping power averaged over the energy of the step, in the detection material and the medium, respectively. Note that since electron steps are usually small, the averaged stopping power can be approximated by the stopping power of the mean kinetic energy in the step. The electron track-ends are calculated as follows:

\[ TE_{\text{det}} = TE_{\text{med}} \left( \frac{\Delta_{\text{coll,det}}}{\Delta_{\text{coll,med}}} \right) \]  

(Note that this scaling of the track-end, represented by the ratio of stopping powers, comes from the analytic definition of track-ends at Equation 4.14. This accounts for the fact that the number of electrons with kinetic energy falling below \( \Delta \) and being absorbed is scaled by that factor. Also note that this approach to stopping-power ratio calculations requires that the choice of the simulation energy threshold must be equal to \( \Delta \).

In each algorithm, the transport calculation is only performed in a single medium (the medium in which the secondary electron fluence has been established) and in which the stopping-power ratio is meant to be calculated. As it is defined by Bragg–Gray cavity theory, the electron fluence is only calculated once and is thus assumed unperturbed by the presence of the detector, and this condition is achieved in both approaches. It is clear that the calculation of stopping-power ratios is based on Spencer–Attix cavity theory and Bragg–Gray conditions, and therefore can only represent realistic detector response in specific conditions. In reality, both conditions of Bragg–Gray are violated. Indeed, the electron fluence is perturbed by the presence of the cavity, and photons interact in the volume of interest, filled with either material. Therefore, perturbation factors must be evaluated to convert the detection cavity response to the medium response accurately (as stated by Equation 4.8). Another matter to consider is the choice of the parameter \( \Delta \), which can lead to contradictions (see, e.g., Bouchard, 2012). Nonetheless, stopping-power ratios are useful to determine the absorbed-dose energy dependence of a given detection material.

### 4.3.2 Interaction Coefficients

Other important quantities of interest that can be calculated with Monte Carlo simulations are mass–energy absorption coefficients. In the context of dosimeter response for in-air and in-phantom measurements, several studies have addressed kilovoltage photon beams (Seuntjens et al., 1987, 1988; Ma and Seuntjens, 1999) as well as high-energy photon beams (Cunningham et al., 1986; Furhang et al., 1995). Calculated data have been used to obtain correction factors in respect to dosimetry protocols (AAPM’s TG-21, 1983; IAEA’s TRS-277, 1987; AAPM’s TG-51, 1999; IAEA’s TRS-398, 1999; AAPM’s TG-61, 2001) as well as detector response scaling. The average mass-absorption coefficients in a given medium is defined as follows:

\[
\left( \frac{\mu_{\text{med}}}{\rho} \right) = \int_{0}^{E_{\text{max}}} \varphi_{\text{f}}(hv)(\mu(hv)/\rho)_{\text{med}} E_{\Delta}(hv)dhv \int_{0}^{E_{\text{max}}} \varphi_{\text{f}}(hv)hvdhv, \quad (4.17)
\]

where \( \varphi_{\text{f}}(hv) \) is the photon fluence differential in energy (in \( \text{MeV}^{-1} \text{ cm}^{-2} \)), \( \mu \) is the energy-dependent linear attenuation coefficient in the medium (in \( \text{cm}^{-1} \)), and \( E_{\Delta}(hv) \) is the average energy being absorbed by the medium from photon interactions with initial energy \( hv \). Note that one needs Monte Carlo simulations to evaluate \( E_{\Delta}(hv) \). The evaluation of such coefficients requires linear attenuation data and a proper knowledge of the photon fluence spectrum, either measured or calculated with Monte Carlo simulations under specific conditions. Systematic compilations of mass–energy absorption coefficients for elements and materials of dosimetric interest were performed by Hubbell (1982) and Seltzer (1993) for monoenergetic photons. To avoid storage of a large number of spectra, direct scoring techniques with Monte Carlo can be used to calculate average ratios under specific conditions.

### 4.4 Modeling Detector Response

#### 4.4.1 Ionization Chamber Correction Factors

Over the last decades, considerable improvements in particle transport accuracy (e.g., Bielajew et al., 1985; Bielajew and Rogers, 1987; Andreo et al., 1993; Bielajew et al., 1994; Baro et al., 1995; Sempau et al., 1997; Kawrakow and Bielajew, 1998a, 1998b; Kawrakow, 2000a; Bote and Salvat, 2008) have allowed the use of Monte Carlo methods for detector response simulation to an accuracy of 0.1% or better (Kawrakow, 2000b; Seuntjens et al., 2002; Poon et al., 2005; Sempau and Andreo, 2006). Such improvements have allowed studies to be performed in dosimetric situations that could not be done by other means with the same level of accuracy. A recent example is the behavior of ionization chamber response in nonstandard beam configurations (small beams or modulated beams), which was extensively studied using Monte Carlo models and compared with experimental measurements (Paskalev et al., 2003; Bouchard and Seuntjens, 2004; Capote et al., 2004; Sánchez-Doblado et al., 2005). These studies demonstrated the need for Monte Carlo calculations in the evaluation of correction factors for such dosimetric situations.

More recent publications used Monte Carlo modeling to perform studies or to generate accurate data related to ionization chamber response to dose in context of standards or reference dosimetry protocols (Buckley and Rogers, 2006a,b; Burns, 2006; Kawrakow, 2006; Mainegra-Hing et al., 2008; McEwen et al., 2008; Wang and Rogers, 2008, 2009a,b,c; Wulff et al., 2008a; Zink and Wulff, 2008, 2011; Bouchard et al., 2009; Crop et al., 2009; Tessier and Kawrakow, 2010).

In general, detector sensitivity (or response) calculations cover three areas of importance: (1) in-phantom reference dosimetry, (2) in-phantom relative dosimetry, and (3) in-air chamber correction factors.
In the context of reference dosimetry, calculation of component-specific and replacement perturbation factors can be performed by taking the ratio of absorbed doses calculated in a series of detector geometries, only differing in their components or in the composition of the cavity. In this manner, perturbation correction factors can be calculated as a ratio of doses scored in geometries connected to each other by chain of correlated geometries. The chain of correlated geometries can consist of a fully-modeled chamber having components gradually removed from it to calculate the effect of a specific perturbation. It can also contain a virtual volume in which dose is scored, such as, for example, a bare chamber scoring volume in which stopping-power ratios are calculated or a small scoring sphere serving to evaluate dose at a point. The ratio of absorbed doses in the user-defined scoring volumes corresponds to a perturbation factor related to the differences in characteristics of the volumes (i.e., their geometry or composition). For example, the ratio between calculated absorbed doses to the chamber without wall and the fully modeled chamber is, by definition in this case, the wall perturbation factor. From this starting point, the ratio between calculated absorbed doses to the chamber without wall and central electrode and the chamber without wall is, also by definition in this case, the central electrode perturbation factor. The calculation chain goes on until the all-scoring volumes are defined to represent the factors in Equation 4.7. This method represents a significant improvement in consistency from approximate calculations or measurements. Figure 4.1 illustrates the principle chain to determine perturbation factors used by Wulff et al. (2008b).

For relative dosimetry, Monte Carlo modeling of ionization chamber response has been applied to aid an accurate conversion of depth ionization into depth dose for both photon and electron beams. In electron beams, this entailed calculation of stopping-power ratios (Burns et al., 1995) and perturbation correction factors as a function of depth (Buckley and Rogers, 2006a,b; Verhaegen et al., 2006). In photon beams, stopping-power ratios have been shown to be quasi-depth independent, except for the build-up region (Andreo and Brahme, 1986) but recent efforts have readdressed the issue of quasi-depth independent, except for the build-up region (Andreo 2006a,b; Verhaegen et al., 2006).

Other detectors have also been modeled in the context of relative dosimetry. Diode detectors are very efficient for electron beam PDD measurements, since at clinically relevant depths when placed with the center of the sensitive volume at the appropriate depth, measured reading does not require any correction. However, because electrons reach sub-MeV kinetic energy in the high-gradient region of the PDD curve beyond $d_{\text{max}}$, and because the relative contribution of bremsstrahlung photons to dose increases for depths beyond $d_{\text{max}}$, one might expect that the response of the diode must be corrected in this dose fall-off region and especially the bremsstrahlung tail. With the effective atomic number of water roughly between $Z = 7$ and $Z = 8$, and the one of amorphous silicon $Z = 14$, the large differences in electron scattering power at low kinetic energies might lead to an appreciable shift in the position of the depth of 50% maximum dose ($R_{50}$) and the practical range ($R_{p}$). However, Wang and Rogers (2007) extensively studied diode response to electron beam with depth based on a complete Monte Carlo model and conclude that the effect on the position of $R_{p}$ is well below 1 mm when using a diode to measure electron depth dose curves. However, they showed that corrections on the depth dose curve in the bremsstrahlung tail amounts to 5–10% of the maximum dose. This has been demonstrated experimentally for 6–21 MeV clinical electron beams, where the dose in the bremsstrahlung tail relative to the maximum dose measured with the TG-51 dosimetry protocol was compared to charge collection ratios with a restricted water-to-detector mass stopping-power ratio correction and highly constrained Monte Carlo simulation, with the parallel plate chamber underresponding in the bremsstrahlung tail by 10%, the silicon diode overresponding by up to 5% (see Faddegon et al., 2005, 2009).

![Figure 4.1](image-url)  
**Figure 4.1** Illustration of the chain technique of Wulff et al. (2008b) for calculating ionization chamber perturbation factors. The various perturbation factors are given by the dose ratios from one step to another in the ionization chambers’ cavity (1–4) and the dose to a small portion of water (5). The step from models 3 and 4 can be further subdivided into separate calculation of $P_{\text{wall}}$ and $P_{\text{diroom}}$ (From Wulff, J., Heverhagen, J. T., and Zink K. 2008b. Monte-Carlo-based perturbation and beam quality correction factors for thimble ionization chambers in high-energy photon beams. *Phys Med Biol* 53: 2823–2836. With permission.)
An important early area of application for chamber simulations was in the context of air kerma radiation standards. A historical application has been the calculation of the wall correction factors in air kerma standards for the dissemination of the air kerma standards through protocols such as TG-21 and TRS-277. In-air ionization chamber wall corrections are calculated in a process of correlated scoring of the energy deposited by (1) the total energy deposition, that is, all secondary electrons and their progeny from primary and scattered photons, (2) the “first-collision” energy deposition by all secondary electrons and their progeny from only primary photons, and (3) the “unattenuated first-collision,” which is the first-collision energy deposition, corrected for attenuation of the primary photon. The wall correction is calculated as a product of two factors, the attenuation correction and a scatter correction. The attenuation correction is defined as a ratio of doses scored in (3) by those in (2), whereas the scatter correction is defined as a ratio of doses scored in (2) by those in (1). All scoring is done simultaneously for each history, making the scoring highly correlated, which greatly reduces the statistical uncertainty in the ratios. Despite rigorous proof of accuracy by Bielajew and colleagues and despite the low uncertainties to which wall corrections could be calculated, it was not until detailed experiments by McCaffrey et al. (2004) and Buermann et al. (2003) confirmed the Monte Carlo correction factors that the standards dosimetry community accepted Monte Carlo-based wall correction factors.

A final application that has received recent attention is the reevaluation of free-air chamber correction factors by Monte Carlo techniques as recently reviewed in detail by Burns and Buermann (2009).

### 4.4.2 Simulation Geometries

In general, common Monte Carlo systems provide a geometry package in which radiation transport is simulated. In the context of photon and electron beam radiation therapy, accurate detector response calculations are performed using mainly two systems: (1) EGSnrc (Kawrakow, 2000a) and (2) PENELOPE (Salvat et al., 2009). Both systems have powerful tools to describe detector geometries. As for EGSnrc, the user code cavity and egs_chamber (Wulff et al., 2008b), being specialized in detector response simulation, allow for the use of several geometries based on the egs++ library (Kawrakow, 2005). Typical user-defined geometry can be done using simple solids such as boxes, spheres, or cylinders along with the use of Boolean operators (union, intersection, etc.) on these objects to combine them logically and make complicated geometries. Also, geometries can be defined by the surfaces surrounding them, using Boolean operators to define volumes from a combination of simple surfaces, for example, planes or infinite cylinders, and solid objects. The structure of the library can also allow the definition of new geometry classes. In PENELOPE, the main idea is the use of quadratic surfaces, defined by a second-order equation with \( x, y, \) and \( z \), to define the geometries. Solid bodies are constructed with a combination of surfaces in which the space is filled with a homogeneous medium. The priorities of the bodies are specified by the user, allowing embedding bodies into other volumes to obtain complex geometries. In both systems, rotations and translations of the geometries are allowed, making it possible to simulate any detector with very good precision in detail. Figure 4.2 shows the chamber modeled by Crop et al. (2009) using the egs++ library.

### 4.4.3 Efficiency Improvements

To reach a precision level comparable to the calculation accuracy (i.e., 0.1% (Kawrakow, 2000a)), simulations must be highly efficient. The first matter being considered in such types of application is the determination of region indexing, which without any proper consideration can become time consuming when the geometry is complex. The design of algorithms allowing efficient transport through boundaries has been given a high priority in the design and implementation of some codes, like the EGSnrc geometry package egs++ (Kawrakow, 2005). Calculations performed in geometries having a single scoring volume allow for significant efficiency improvements as compared to brute force calculations. One common variance reduction technique (VRT) used in practice for detector response simulation is range rejection (RR), being used in both EGSnrc and PENELOPE. The idea in this technique is that any charged particle having a range

![FIGURE 4.2 Geometrical models of the following chambers achieved by Crop et al. (2009): (a) NE2571, (b) PinPoint 31006, and (c) PinPoint 31016 (images are not on the same scale). (From Crop, F. et al., 2009. The influence of small field sizes, penumbra, spot size and measurement depth on perturbation factors for microionization chambers. Phys Med Biol 54: 2951–2969. With permission.)](image-url)
smaller than its distance to the scoring volume will be discarded. This avoids simulating particle transport that will yield no effective result. One approximation in this technique is in the amount of bremsstrahlung radiation that will reach the scoring volume. To control this approximation, the user can define a threshold below which RR is considered.

Correlated sampling (CS) is also a VRT having good affinities with detector simulation. For applications where relative detector response is assessed, some information on particle transport being common to several geometries can be recycled. For example, if one is interested in calculating the wall correction factor, two geometries are defined at the same position: (1) a fully modeled chamber and (2) a chamber modeled without wall. It is clear that for any particle being transported from the beam in either geometry, the transport will be identical until the particle reaches the region where the wall is located. Defining a CS region enclosing both detectors, one can save a significant amount of computational time simulating both geometries using the same information gathered on the surface defining the CS region. Another example is for detector response relative to another location, to mimic IC measurements. In such application, the CS region is made large enough to enclose all detector positions. When scoring with the CS technique, the energies and variances are not only scored but the covariances as scored energies are also positively correlated between each other. Mathematically, this implies that the uncertainty on the ratio is smaller than the one obtained for independent simulations using the same number of histories. The CS technique was implemented in CSnrc by Buckley et al. (2004), a modification of the user code CAVRZnrc, and is illustrated in Figure 4.3. The user code egsc_hamber, being a modification of the code cavity, was designed by Wulff et al. (2008a), defining the intermediate phase-space scoring (IPSS) technique to use CS in the subgeometries. Finally, the user code CS_cavity, also being a modification of the code cavity, was implemented by Crop et al. (2009). While in early studies, typical gains in efficiency with the CS technique were reported between 10 and 100 (Ma and Nahum, 1993), they are roughly between 20 and 40 for CSnrc and CS_cavity, respectively.

Another powerful VRT relevant to detector response simulation in photon beams is photon cross section enhancement (XCSE). Defining a region around the scoring volume in which the photon interaction cross sections are enhanced, the shower of particles is amplified. The resulting information extracted from this shower, initiated from single particles entering that region, is such that the overall simulation efficiency is optimized. There are strict rules to obey to assure an unbiased result and equal particle weighting. For that matter, particle splitting and Russian roulette are played in such a way that each particle type carry a unique statistical weight, given a XCSE factor (see Wulff et al., 2008b). Note that electrons surviving the range rejection game make an exception to this rule. Figure 4.4 illustrates the combination of CS technique, through the IPSS method, and XCSE in the user code egsc_hamber (Wulff et al., 2008b). Note that during the validation of this code, the gain in efficiency by the use of XCSE itself was reported to be 130 for the cases simulated (Wulff et al., 2008b). When comparing a simulation without any VRT with an optimal simulation (i.e., using RR, CS, and XCSE), total gains in efficiency are between 650 and nearly $10^4$, depending on the nature of the simulation.

4.4.4 Positioning Uncertainties in Detector Response Calculations

In radiation dosimetry, another quantity of interest is experimental uncertainty, which is typically evaluated either by a statistical analysis of observations or by other nonstatistical methods (ISO, 1995). Among several sources of uncertainty, one that could potentially be evaluated with Monte Carlo calculations is the uncertainty on detector response induced by setup positioning errors. This topic is relevant particularly in situations where high accuracy is required and dose gradients are significant, such as nonstandard photon beams (Chung et al., 2010) or electron beam dosimetry (Zink and Wulff, 2009). For these applications, the effect of positioning errors on detector response can be considerable and have a substantial effect on the confidence of the measured data.

In a recent paper, Bouchard et al. (2011) constructed and implemented a method in the user code egsc_hamber to evaluate these effects in a systematic manner. The method is based on...
on the principle that Monte Carlo simulations of radiation transport can be used to score other quantities than typical energy deposition. In fact, any Monte Carlo method consists of scoring statistical estimators along with their respective uncertainties. In the present context, the quantity of interest is the standard deviation of the detector response, each individual response calculated in a randomly generated geometry, based on a user-defined statistical behavior of geometry positioning. Intuitively, the simplest approach to calculate such quantity would be to proceed as follows. First, one generates random detectors or geometry positions (including translations and rotations), this in respect to what is defined as probable positioning errors (according to certain probability distributions). Second, one calculates the dose response of the detector at each position. Finally, one estimates the standard deviation of the calculated responses and normalizes it to the mean response. While this seems to be a trivial way to proceed, the direct result from this intuitive approach is statistically biased, since results from Monte Carlo simulations are estimated and contain statistical uncertainties (Bouchard et al., 2011).

In the method proposed by Bouchard et al. (2011), an unbiased estimator of dose uncertainty induced by positioning errors is constructed for accurate calculations. An unbiased estimator of dose ratios, as is used in the context of CS, is also built taking into account the statistical correlations. The method is implemented in egsc_hamber and allows for positioning errors on (1) the phantom according to the beam and (2) the detector according to the phantom. Several spatial operations (i.e., translation and rotation) as well as probability distributions (i.e., Gaussian and uniform) are allowed to characterize the random nature of positioning errors. Although the method requires an optimization procedure to obtain efficient calculations, it is shown that with optimal parameters one can achieve the same statistics as for standard dose calculations with only twice the number of histories. In typical situations, since the statistics on the uncertainty estimators need to be typically of the order of a few tenth of a percent, calculations can be time consuming. Nonetheless, such simulations find uses in situations where no other means to evaluate these effects exist, especially in nonstandard fields. Figure 4.5 shows results obtained by the method by Bouchard et al. (2011) for small field output factor measurements using a diode.

4.5 Accuracy of Detector Response Calculations

4.5.1 General Considerations

In assessing the accuracy of Monte Carlo calculated detector response, there are four main sources of systematic uncertainties:

1. The accuracy of the CH transport system
2. The accuracy of the cross-section data in the system, including both fundamental properties of elements (Z-values, etc.) and how well the true composition of the materials of the real detector are modeled (impurities, etc.)
3. The accuracy of the geometry of the model of the detector
4. The accuracy of the radiation source model (i.e., small field, collimation system, spot size, etc.)

4.5.2 Accurate Condensed History Technique

During the simulation of charged particle transport, the analog simulation of discrete interactions is limited to a certain
energy threshold below which the interactions are condensed into a single step, or CH (Berger, 1963). The calculation of electron (and positron) condensed steps is performed based on the continuous slowing down approximation (CSDA) and the multiple scattering theory. It is worth noting that both approaches must, on average, converge to an accurate result. During electron transport with the CH technique, the two main algorithms involved are (1) the electron-step algorithm, where the electron nonlinear trajectory is determined, and (2) the boundary-crossing algorithm. The approach by Kawrakow (2000a), implemented in the EGSnrc system, uses several improvements of the techniques behind these algorithms, as compared to earlier versions of the system (Nelson et al., 1985; Bielajew et al., 1994). The constraints of the Molière multiple scattering theory were resolved by the implementation of the Goudsmit–Saunderson formulation (Kawrakow and Bielajew, 1998a), diminishing the truncation error by two orders of magnitudes (Kawrakow and Bielajew, 1998b). The accuracy of the boundary-crossing algorithm was also improved by the use of a single-scattering mode near the boundaries, resolving a known singularity (Kawrakow, 2000a). To evaluate the potential accuracy of Monte Carlo simulations in radiation dosimetry and detector simulation applications, the convergence of these improved algorithms was extensively studied.

4.5.3 Fano Test

Accurate Monte Carlo calculation of the extrinsic response of gas-filled ionization chambers or cavities is one of the most stringent tests of a CH Monte Carlo code. The electron transport physics, the boundary-crossing algorithm, the evaluation of energy loss along the CH step, as well as the cross sections must be accurate. During the 1980s and 1990s, the ability of Monte Carlo to accurately calculate chamber response was assessed by comparing results to a calculation of the so-called Fano test. This test relies on the validity of the Fano theorem (Fano, 1954), which states that under conditions of CPE, the medium electron fluence differential in energy and angle is independent of density variations from point to point. The theorem can be used to test the extent to which the electron step algorithm, path length correction, boundary-crossing algorithm, and energy loss evaluation have been implemented adequately in a Monte Carlo simulation program when dealing with transport in a situation consisting of multiple regions and densities.

The Fano test is usually carried out on a cavity filled with the gaseous form of the phantom material in which the cavity is embedded. The idea is that the gaseous form of the wall material has exactly the same cross sections as the wall material itself but has a 1000-fold lower density. This means that density effect as part of the stopping power of the cavity gas and wall must be taken as the same or artificially removed from the calculation. The Fano test first requires to setup, artificially, perfect CPE. There are two vintages of the test, differing in the way this is achieved.

One way is close to the physical reality of a photon beam interacting in a material and effectively present itself as a distributed source of charged particles set in motion. In a typical cavity (ionization chamber) setup, the beam is interacting in the phantom (or wall) of the simulated ionization chamber and either restore the primary photon once it has interacted, discard the scattered photon, and follow the secondary electron slowing down (the regeneration technique) or, alternatively, by dividing the cavity dose by the in-air correction factor for photon attenuation and scatter. Effectively, all this means that while the photon beam interacts, it is not allowed to be scattered and scattering and attenuation effects are “undone.” This clearly represents a situation where energy is not conserved; in fact, the radiant energy inserted in the form of a photon beam is recovered after the secondary electrons are generated and the photon beam is used for no more than generating a distributed electron source in the medium. The advantage of performing the test in this manner, however, is that the energy inserted in the medium can be retrieved from photon mass energy transfer coefficients.
The Fano test entails the comparison of collision kerma in the wall, under CPE with dose to the cavity, as follows:

\[ D_{\text{cav}}^{\text{wall}} = \frac{D_{\text{cav}}}{A_{\text{wall}}} = K_{\text{coll,wall}}. \] (4.18)

This technique has been used by several authors during the investigation of accuracy of the EGSnrc code Kawrakow (2000b), Seuntjens et al. (2002), the PENELLOPE code (Yi et al., 2006), and the GEANT4 code (Poon et al., 2005).

The Fano test requires the calculation of collision kerma to the wall (or build-up medium), which requires the estimation of energy fraction expended in bremsstrahlung in the wall \(g_{\text{wall}}\). This is usually done in auxiliary calculations and should be based on the same photon interaction data as used in the cavity dose calculation so that the evaluation of Equation 4.18 becomes cross-section independent.

A second and more direct way, which avoids these auxiliary calculations, is to realize Fano conditions by directly initiating electron fluence with uniform intensity per unit mass in an infinite medium with a gap of infinite extent. In this case, the dose is just the energy of the electrons times their number per unit mass. Sempau and Andreo (2006) used this method to compare dose calculations with the CH Monte Carlo simulations with event-by-event calculations and arrived at a consistency of 0.2% with PENELLOPE.

### 4.5.4 Accuracy of Physical Data

The Fano test does not assess one of the components in an accurate detector response calculation in realistic situations, that is, the accuracy of the interaction data. There are several sources of cross-section data in literature, as well as several models behind them, and some physical data depend on few parameters, such as electron and positron stopping powers, which depend on the mean excitation energy \(I\)-value \(I\text{-value) (ICRU37, 2007). A recent study by Wulf et al. (2010) studied systematic variations of ionization chamber quality correction factors \(k_q\) induced by systematic changes in cross-section data and \(I\)-values. It was found in general that the effect on \(k_q\) factors by uncertainties in photon cross-section data, being estimated to be of the order of 1–2%, is only up to 0.4% for typical thimble chambers. As for the effect of uncertainties on the \(I\)-value, the effect is remarkable only for \(Z = 6\), assuming an uncertainty on the \(I\)-value of 6.1% for that material, yielding an effect of up to 0.37% on \(k_q\) factor calculations. It is worth noting that another similar study by Muir and Rogers (2010) comes to slightly different conclusions, obtaining systematic errors up to 0.55% for \(k_q\) from uncertainties in photon cross sections, and up to 0.19% on \(k_q\) from uncertainties on \(I\)-values (here, an uncertainty on the \(I\)-value for \(Z = 6\) was taken to be 4.5%).

The value of the energy necessary to create an ion pair in air, \(W_{\text{air}}\), has been the subject of considerable recent research and controversy. The complication is that most sources of values of \(W_{\text{air}}\) involve graphite cavity ionization chambers in conjunction with graphite calorimeters and the value of \(I\) for graphite enters in the determination of \(W_{\text{air}}\), through the restricted collision stopping-power ratio graphite to air. Büermann and Hilgers (2007) compared exposure measurements at 300 and 400 keV using a large free-air and a graphite cavity chamber. This allowed a determination of the graphite-to-air mass stopping power ratio to a stated uncertainty of 0.3% and the \(I\)-value for graphite consistent with this is close to the Bichsel and Hiraoka (1992) \(I\)-value. Büermann et al. (2007) did a second experiment, where they use a graphite cavity chamber in the physikalisch technische bundesanstalt (PTB) water calorimeter in a \(^{60}\text{Co}\) beam, and used a Monte Carlo calculated ratio of dose to water to dose to cavity to compare absorbed dose from the ionometric to the calorimetric method. Reconciling both methods for absorbed dose to water determination requires either an increase in \(I\)-value to 87 eV or a 1.5% reduction in the \(W_{\text{air}}\) value (or a combination of both). A result qualitatively similar, albeit only amounting to an effect of 0.7%, is obtained by considering the key comparison data of worldwide absorbed dose standards at \(^{60}\text{Co}\), maintained by the bureau international des poids et mesures (BIPM).

Thomson and Rogers (2010) reanalyzed the calorimeter experiments that led to the determination of the product of the energy necessary to create an ion pair in air per charge, \(W_{\text{air}}\), with the graphite-to-air mass stopping power ratio. This was done by reanalyzing the gap corrections in the graphite calorimeter experiment and by analyzing the sensitivity of the product to secondary particle production thresholds and \(I\)-values. As suggested, extracted \(W_{\text{air}}\) value from this analysis is 33.68 J/C \(I\pm 0.2\%), which is 0.8% lower than the currently accepted 33.97 J/C \(I\pm 0.1\%) (Boutillon and Perrorche-Roux, 1987). The impact of the uncertainty in \(W_{\text{air}}\) on \(k_q\) was studied by Muir and Rogers (2010). They demonstrated an effect of 0.5% on \(k_q\) factor calculations based on the uncertainty of 0.75% on \(W_{\text{air}}\) for air. In a more recent paper, Muir et al. (2011) showed an excellent agreement between measured and Monte Carlo calculated \(k_q\) factors. From their study, they concluded that an upper limit of variation of 0.4% exists on \(W_{\text{air}}\) in the energy range from cobalt-60 to 25 MV.

### 4.5.5 Effects of Machine Modeling

In the context of detector response calculation, machine modeling is a crucial step of the procedure. While it can be understood from Equation 4.8 that detector response is dependent on the electron spectrum, stopping-power ratios are not quite sensitive to the beam lateral fluence (Sánchez-Doblado et al., 2003; Bouchard et al., 2009). Among the parameters determining the model, the details in the geometry and materials and details on the electron beam, such as the radial distribution and the energy, are important parameters to consider. A paper by Sheikh-Bagheri and Rogers (2002) is a good starting point to systematically determine these parameters from experimental measurements. Based on a proper model of the machine geometry from the manufacturer’s specifications, one can estimate the incident electron beam energy and spatial distribution, calculated off-axis factors and central-axis depth dose curves can be compared with measurements. Central-axis depth dose data is useful to determine
the mean energy of the electrons. Once the mean energy is determined, the in-air off-axis factors are used to derive the spot size of the electron beam spatial distribution.

Recent studies reported that the detector response to linac beams is significantly sensitive to the spot size of the incident electron (Francescon et al., 2008; Crop et al., 2009). During the simulation of small beam output factors, Francescon et al. (2008) reported that a change of 1.2 mm in the full width half maximum (FWHM) can have an effect up to 6% on the calculation of output factors with several detectors, such as ionization chambers, diodes, and diamond detectors. Another study by Crop et al. (2009) concludes that the strong influence of the spot size on ionization chamber response for small field sizes limit their use. The difficulties related to the estimation of the spot size depend on how the machine is modeled from the start. If one adjusts the FWHM based on open field measurements, such as recommended by Sheikh-Bagheri and Rogers (2002), small errors in the shape of the flattening filter can potentially yield erroneous spot sizes. One way to improve this method is to determine the FWHM from small field output factor measurements and explicitly accounting for the detector in the model (Francescon et al., 2008; Kawrakow, 2011). Then the shape of the flattening filter and other influencing components of the linac are adjusted with large field profiles. This method has demonstrated to improve the agreement between the calculations from the model and experimental measurements (Francescon et al., 2011). Another approach is to measure the spot size and other simulation details directly, putting narrow constraints on the parameter adjustment (Sawkey and Faddegon, 2009).

4.5.6 Size of Volume of Interest

The size of the volume in which absorbed dose is calculated influences the calculation accuracy. While the Monte Carlo method intrinsically defines energy thresholds below which particles deposit their energy locally, this approximation can break down for volumes with dimensions comparable to the range of electrons having sub-threshold energies. As energy is being conserved during particle transport simulation, what contributes to the error between the dose calculated with model and the true dose is the fraction of energy being locally absorbed that could be transported across the boundaries of the volume of interest. Therefore, the range (of mean free path) of the particles having kinetic energy (or energy) below the threshold determines the critical region of the volume where the absorbed energy is approximated. If this region is small compared to the volume, the error is also small. Therefore, an appropriate choice of energy thresholds is necessary to calculate dose accurately in small volumes.

In typical cases, the accuracy of the method related to the size of the scoring volume can be evaluated by showing the convergence of the calculations as a function of the energy thresholds. However, it can be argued that there are some exceptions in which the particle range or mean free path do not strictly decrease with decreasing energy. One simple example is the behavior of photon cross sections with energies around different atomic shells, where discontinuities are observed due to a drastic change in the availability of electrons. However, in general, the particle range (or mean free path) decreases with decreasing energy threshold. Therefore it is expected that the smaller is the scoring volume, the smaller the energy threshold needs to be for accurate calculation. However, it is also worth noting that since the Monte Carlo method is a semiclassical approach, it is limited to the transport of particles with kinetic energy (or energy) equal or larger than 1 keV, by Heisenberg’s uncertainty principle (Thomson and Kawrakow 2011). This implies that the accuracy of the method is seriously compromised in the case where the range (or mean free path) of particles having a kinetic energy (or energy) of 1 keV is comparable to the size of the cavity.

4.5.7 Effects of Geometry Accuracy

The accuracy to which geometrical details are described in detector models also has an effect on the calculation accuracy. The achievement of such detailed modeling depends on the tools available, which have considerably evolved over the past decade. While earlier studies modeled ionization chambers with simplistic geometries (Bouchard and Seuntjens, 2004; Capote et al., 2004; Sánchez-Doblado et al., 2005), the accuracy to which they generate ionization chamber response-related data was limited. With the new egS++ library (Kawrakow, 2005) and the geometry package in PENELOPE (Salvat et al., 2009), detector geometry can be fully modeled. A study by Ubrich et al. (2008) demonstrated that ionization chamber components can have a crucial effect on their response, specifically for photon beams in the kilovoltage range. Indeed, they showed that a small difference in the shape of the central electrode results in a deviation of 8.5% for beam quality correction factor $k_Q$. However, for megavoltage photon beams, recent studies demonstrated that the effects are small (i.e., a few tenth of a percent) in reference conditions (Muir and Rogers, 2010; Wulff et al., 2010). Figure 4.6 illustrates the method of Muir and Rogers (2010) that showed maximum differences in $k_Q$ factors of 0.2% for the geometries used: (a) fully-modeled chamber, (b) chamber with simplistic stem, and (c) cylindrical wall with central electrode. From these reported studies, it can be concluded that although for typical situations the geometrical accuracy has a small impact on calculations, they remain an important matter to be considered, especially for nonstandard beam configurations. This is relevant especially for kilovoltage beams, metallic detector components, and nonstandard megavoltage beams. In fact, photoelectric effects can affect detector response significantly, and beams where lateral disequilibrium exists can have an important effect on the modeling of volume averaging effects and the displacement of the water volume by the air cavity.

4.6 Small and Nonstandard Field Dosimetry

4.6.1 Small Fields

Small fields and combinations of small fields are increasingly used in radiation therapy and can be generally termed nonstandard...
fields as they are different from conventional flat radiation therapy beams. For the purpose of measurement dosimetry, there are three characteristics that determine the presence of small field conditions in external photon beams. These are: (i) loss of lateral CPE, (ii) partial occlusion of the primary photon source by the collimating devices, and (iii) size of the detector compared to the beam dimensions. The first two characteristics are beam related, while the third is detector related and depends on the relative size of the detector versus the size of the radiation field.

Loss of lateral CPE occurs in photon beams if the beam radius becomes smaller than the maximum range of secondary electrons. Monte Carlo calculations by Li et al. (1995) have aided in quantitatively establishing a criterion for small field conditions by determining the minimal radius of a circular photon field for which collision kerma, corrected for center of electron production, becomes equal to absorbed dose. The second characteristic is related to the finite size of the primary photon beam source, the extended focal spot, which is usually determined by the FWHM of the bremsstrahlung photon fluence distribution exiting the target. Shielding of the primary source will strongly affect the output of the beam. Monte Carlo calculations involving detailed modeling of the treatment head with special collimators confirm these characteristics (Paskalev et al., 2003; Sham et al., 2008). The third feature that characterizes a small field is the relative size of the detector in comparison to the size of the radiation field. Finite-size detectors will produce a signal that is proportional to the average dose over the sensitive volume of the detector and this is affected by homogeneity (or lack thereof) of dose over the detection volume (volume averaging) in a small field; from this signal, a deconvolution process is required to derive the dose at a point. The ideal detector for small fields samples the dose at a point, is water-equivalent, and has a linear dose response, which is energy independent and dose rate independent. The minimal chamber size, however, is determined by technical considerations related to the magnitude of the extracameral signal as compared to the signal from the cavity. Since detector perturbations such as volume averaging and density effects are dramatically influenced by the presence of high gradients and charged particle disequilibrium, small field conditions are also assumed to exist when the sensitive volume of the detector is less than a lateral equilibrium length away from the edge of the field.

In a recent proposal by a consultancy group of the International Atomic Energy Agency (IAEA) in collaboration with the American Association of Physicists in Medicine (AAPM), Alfonso et al. (2008) proposed the concept of an intermediate machine-specific reference (msr) field to link reference dosimetry in small fields to the calibration chain in conventional reference fields. The reason for this proposal is that many of the specialized external beam machines are unable to meet field size and reference dosimetry setup requirements of traditional absorbed dose protocols and one could, therefore, not follow the reference dosimetry protocol recommendations of TG-51 or TRS-398, leading to clinical dosimetry uncertainties. Alfonso’s proposal gives a framework for calibration in msr fields and output measurements in other fields relative to msr fields. The proposal also defines correction factors on output factor measurements taking into account the nature of the detector used for the output factor measurements. One can easily show that the ratio of correction factors for a pair of detectors of different type is equal to the ratio of the readings of that pair of detectors. Sets of ratios of these correction factors for distinct detector types can thus be extracted from measurements for varying field size. Monte Carlo calculations have been used extensively to estimate correction factors for reference dosimetry or to correct measured output factors in small field radiation therapy and can, thus, be used to verify the above-mentioned ratios. Two recent example studies that have followed the Alfonso nomenclature are the extensive small field dosimetry work by the Francescon group (Francescon et al., 2008, 2011) and Cranmer-Sargison et al. (2011). Francescon et al. (2011) performed a detailed analysis on the correction factors, \( k_{\text{msr}} \), for different small field detectors and field sizes on two accelerators. What makes the work specifically useful is that a sensitivity investigation of the effect of source parameters on the uncertainties is included. Similarly, Cranmer-Sargison et al. (2011) recently published values of \( k_{\text{msr}} \) for a set of diode detectors and different field sizes. They discuss in detail the Monte Carlo commissioning of the source and the coupling of this to the detailed modeling of the detectors and conclude that the work provides a basis for correcting measured output ratios for diodes.
4.6.2 Intensity-Modulated Beams

Intensity modulated radiation therapy (IMRT) photon beams, a second form of a nonstandard field, is another type of delivery not complying with absorbed dose reference dosimetry protocols. Earlier studies have demonstrated that additional correction factors may have to be considered for accurate IMRT field calibration (Bouchard and Seuntjens, 2004; Capote et al., 2004). While stopping-power ratios were demonstrated to have a small sensitivity to such configurations (Sánchez-Doblado et al., 2003), ionization chamber gradient effects were shown to be the major factor responsible for the significance of these corrections (Bouchard et al., 2009). In a second part of the Alfonso et al. (2008) proposal, the term plan class specific reference field (pcsr) is coined to define a calibration field representative of a modulated delivery for a class of IMRT treatments. The formalism defines the pcsr correction factor to correct the calibration coefficient obtained in conventional reference conditions for conditions occurring in dynamic nonstandard fields. Alfonso et al. (2008) recommend that the factors are calculated with Monte Carlo simulation alone or measured using a suitable detector applying corrections obtained from a Monte Carlo simulation. With tools such as EGSnrc (including BEAMnrc (Rogers et al., 2009)) or PENELOPE, detailed models of modulated fields by dynamic multileaf collimators and detector correction factors have already been established for composite deliveries using step-and-shoot (Sánchez-Doblado et al., 2005) and dynamic techniques (Chung et al., 2010; De Ost et al., 2011) and will allow detailed quantification of pcsr correction factors under different delivery scenarios. One of the issues with the implementation of the pcsr concept is the criterion imposed on such fields with respect to reproducibility, modulation, and homogeneity of the dose distribution. Monte Carlo calculations and measurements such as by Chung et al. (2012) will help establish suitable criteria for defining and clinically implementing these fields.

4.7 Conclusions

Monte Carlo calculations have been part of measurement dosimetry developments since the late 1970s with applications ranging from stopping-power ratios, wall correction factors to absolute calibration (Bouchard and Seuntjens, 2004; Capote et al., 2004). As electric field in ionization chamber dosimetry, the in silico design of novel detectors with superior dosimetric characteristics, or the on-the-fly Monte Carlo simulation of detector response as part of a Monte Carlo treatment planning system commissioning, to name a few. Finally, Monte Carlo calculations of ionization chambers have enabled complex studies of modulated field dosimetry, which will also lead to more accurate, better tuned, and conceptually simpler clinical reference dosimetry protocols.

References


Francescon, P., Cora, S., and Satariano, N. 2011. Calculation of $k(Q_{\text{clin}}, Q_{\text{msr}}) \times (f_{\text{clin}}, f_{\text{msr}})$ for several small detectors and for two linear accelerators using Monte Carlo simulations. *Med Phys* 38: 6513–6527.


5.1 Introduction

It is well accepted that Monte Carlo simulation methods offer the most powerful tool for modeling and analyzing radiation transport for radiotherapy applications. One of the most frequent and important uses of Monte Carlo modeling in external beam radiotherapy is the creation of a virtual model of the radiation source. Applications are source design and optimization, studying radiation detector response, and treatment planning. Since the vast majority of radiotherapy is performed with megavolt (MV) photon beams from a linear accelerator (linac), this is also reflected in the research efforts. To a lesser extent, external beam radiotherapy is performed with electron beams, lower-energy photon beams emanating from an x-ray tube, or hadron beams. This chapter will cover the Monte Carlo modeling of photon beams and Chapter 6 will cover the Monte Carlo modeling of electron beams, while other chapters will deal with hadron beams.

Historically, Monte Carlo methods provided a simple alternative to measurements to derive photon spectra from radiation sources (McCall et al., 1978; Patau et al., 1978; Faddegon et al., 1990, 1991). Monte Carlo techniques were used in the early days of linac design in the 1970s to aid in the optimization of the photon beams (McCall et al., 1978). The first simplified models of photon beams from linacs were reported from the late 1970s till the early 1990s (Patau et al., 1978; Nilsson and Brahme, 1981; Mohan et al., 1985). For linac electron beams, an important early effort was the work by Udale (1988, 1992) in the late 1980s. They introduced the modular approach for Monte Carlo modeling of electron beams, which is nowadays very popular since the advent of the BEAM Monte Carlo user interface (Rogers et al., 1995). Extensive literature reviews of Monte Carlo modeling of electron beams were provided by Ma and Jiang (1999) and of photon beams by Verhaegen and Seuntjens (2003). They cover many details and are in many respects still current.

5.2 Photon Beams from Clinical Linacs

Nowadays, clinically used photon beams in radiotherapy usually are within the energy range of 4–25 MeV. A linac has an essential modular construction as shown generically in Figure 5.1. Different manufacturers may have the components of a linac in a different order but the ones indicated in Figure 5.1 are usually those that are required in a Monte Carlo model. For an overview of linac technology, the reader is referred to several excellent texts available (Greene and Williams, 1997; Van Dyk, 2005). The Monte Carlo codes employed in this chapter are mostly EGS4 (Nelson et al., 1985) or EGSnrc (Kawrakow, 2000) (often in conjunction with the user interface BEAM (Rogers et al., 1995)), GEANT4 (Agostinelli et al., 2003), MCNP (Briesmeister, 2000), and PENELOPE (Sempau et al., 1997). They are all discussed elsewhere in this volume. Techniques to speed up simulations, variance reduction techniques, are also discussed elsewhere in this volume. They are heavily relied upon in linac simulations, which can be prohibitively long, even with today’s powerful computers. A discussion of variance reduction techniques as they pertain to photon linac modeling can be found in the literature (Verhaegen and Seuntjens, 2003).
5.2.1 Components of a Monte Carlo Model of a Linac Photon Beam

The components depicted in Figure 5.1 must be known in detail to build a faithful model of a linac. The composition of materials and alloys; their mass densities; the position, dimensions, and shape of defining surfaces of the components; and their motion must all be known in great detail to build an accurate Monte Carlo model. Knowledge of tolerances may help to determine uncertainties in the calculated results. While seemingly trivial, this is a task that has daunted many workers. It usually requires interaction with linac manufacturers to disclose constructional details. Needless to say, errors made at this stage will often translate into systematic errors in the calculated output of the linac. Therefore, it is of utmost importance that linac blueprints are verified as much as possible and that Monte Carlo linac models are validated against an extensive set of dose measurements of, as a minimum, depth and lateral dose profiles in water, and output factor in water (ratio of dose at a depth in water at the central axis for a certain field size divided by the same for a reference field size). The validation procedure may comprise comparisons against measurements at various source-to-surface distances of a phantom, should treatment planning be the main application. The literature contains many examples of extensive comparisons of Monte Carlo linac models against measurements (Chaney et al., 1994; Mazurier et al., 1999; Deng et al., 2000; Hartmann Siantar et al., 2001; Ding, 2002a).

In 1995, the BEAM/EGS4 user interface was released (Rogers et al., 1995), since then upgraded to BEAM/EGSnrc (or BEAMnrc). This user interface allows easy modeling of radiotherapy linacs and led to a spate of papers presenting photon and electron beam models. It was, and still is, seen as a major step forward in the field. It is currently the most widely used linac simulation package. Figure 5.2 shows a typical example of a simple photon linac model where linac components and particle tracks can be discerned (Figure 5.2a) and where interactions in the flattening filter can be observed in detail (Figure 5.2b). The BEAM code allows easy assembly of a linac model (it also enables building models for x-ray tubes and a few other geometries) with a wide choice of building blocks consisting of geometrical shapes such as disks, cones, parallelepipeds, trapezoids, and so on. The code relies heavily on the fact that linac components do not overlap in the beam direction, which is mostly the case. Separate parts of a linac can, therefore, be built, validated, and studied separately. In addition to linac components that may be defined with any material (the BEAM package provides an extensive cross-section dataset), a wide choice of primary source

---

**FIGURE 5.1** Schematic representation of linac components in a Monte Carlo model of a photon linac. Different linac manufacturers may have the linac components in a different order.

**FIGURE 5.2** (See color insert.) (a) Simplified BEAM Monte Carlo model of an 18 MV linac photon beam. From right to left, we encounter the primary electron beam hitting the target (T), the primary collimator (PC), the flattening filter (FF), and some secondary collimators (SC). Photon tracks are shown in yellow, electron tracks in blue, and positron tracks in red. In (b), a close-up of the particle interactions in the FF is shown in exquisite detail. By depicting the FF in a semitransparent fashion, the details of the geometry (notice the cone inside a cone structure) and the interactions are revealed.
geometries is available. Particle transport can be done for electrons, photons, and positrons. Particles can be tagged according to interaction types, interaction sites, and so on, which provide a powerful beam analysis method.

Other user interfaces similarly intended to enable easy assembly of complex geometries have been introduced, but these have yet to equal the success of BEAM. An example is PENLINAC for the PENELOPE Monte Carlo code (Rodriguez, 2008). The GATE user interface (Jan et al., 2004) for the GEANT4 Monte Carlo code also seems to hold some promise.

Around the same time BEAM was first released, Lovelock et al. (1995) also presented a modular Monte Carlo approach to simulate linac photon beams. Both are predated by papers by Udale (1988, 1992) who introduced modular Monte Carlo models for electron beams. In Lovelock et al. (1995), the radius of a uniform primary electron beam hitting the target and its energy distribution (a truncated Gaussian) could be taken into account. Information on particles crossing a predefined plane could be stored in a phase-space file. Simulations were done in stages, that is, phase-space information from one stage was used as input source for a subsequent stage. Particles could be tagged according to different types of interactions they underwent. They noticed that calculated depth dose profiles in water are relatively insensitive to the primary electron energy, whereas the horns in the lateral dose profiles are a good indicator for the primary electron energy. A lower primary electron energy will result in more pronounced horns. As such, these authors were among the first to introduce a tuning procedure of the Monte Carlo primary electron source characteristics to derive the primary electron energy. This is often needed in case the manufacturer does not provide the required information or where simulations of the electron trajectories in the flight tube are not available. This information is rarely at hand.

The photon linac components that are the most important in the model in Figure 5.1 are the target, flattening filter, secondary collimators such as jaws and multileaf collimator (T, FF, SC), and wedges. After the accelerated primary electron beam exits the flight tube, it possesses a narrow-energy, angular and spatial distribution. This electron beam will hit the target, commonly consisting of a high-Z metal in which the electrons will produce bremsstrahlung photons. It has to be pointed out that in the Monte Carlo models of clinical linacs, usually no modeling is done of the electron beam, prior to exiting the flight tube. This will inevitably entail making assumptions about the primary electron beam, which will be discussed later. The bremsstrahlung photons are then collimated initially by a primary collimator and the photon fluence is differentially attenuated by the flattening filter to produce a reasonably flat dose distribution in water at a certain depth. Target and flattening filters are the most important sources of contaminant electrons, unless a wedge is present. In the generic model of Figure 5.1, the next components to be considered are the monitor ion chamber and the field mirror. Both present only a small attenuation to the photon beam and are often omitted from Monte Carlo models. However, when backscatter to the monitor chamber is the subject of study (Section 5.2.1.3), some model of the ion chamber should be present. The photon beam is then finally shaped and modulated by secondary collimators and beam modifiers such as jaws, blocks, multileaf collimators (MLC), and wedges. In the following sections, Monte Carlo modeling of the various linac components will be discussed in more detail.

5.2.1.1 Primary Electron Beam Distribution and Photon Target

5.2.1.1.1 Photon Target Simulation for Radiotherapy Treatment

The bremsstrahlung photons produced in the target by the primary electrons in the energy range 4–25 MeV are mostly derived from a relatively thin layer on the upstream side. This is because of the quick energy degradation of the electron energy in a high-Z material, the almost linear dependence of the bremsstrahlung cross section on the electron kinetic energy for high energies and electron scattering in the target. At high electron energies, the average bremsstrahlung photon emission angle is given approximately by $m_0c^2/E_n$ ($m_0c^2$ is the electron’s rest energy; $E_n$ its total energy) yielding a strongly forward-peaked angular distribution. Targets in clinical linacs are usually thick enough to stop the primary electrons completely and are for that reason often referred to as “thick targets.” In such targets, the bremsstrahlung photon angular distribution will be spread out because of electrons undergoing multiple scattering, but the result will nevertheless be a strongly anisotropic photon fluence, while the photon spectrum is fairly isotropic. Owing to this complex picture, it is very difficult to obtain the spectral and angular bremsstrahlung photon distribution from thick targets using analytical methods such as the Schiff theory (Koch and Motz, 1959; Desobry and Broyer, 1991). More complex analytical calculation schemes have been developed but the most comprehensive method to generate bremsstrahlung photon distributions from targets in clinical linacs is Monte Carlo simulation. In what follows, we will discuss a few studies related to this topic.

Patau et al. (1978) were among the first to present a simple Monte Carlo model, using an in-house code, of a complete photon beam linac. It consisted of a 5.7 MeV electron pencil beam hitting a W/Cu target, followed by a flattening filter of Pb and an unspecified collimator. The authors calculated photon energy fluences, photon transmission through diverse materials, and electron spectra in water. McCall et al. (1978) were among the first to attempt to improve the design of a linac by performing Monte Carlo simulations. They used the EGS3 code (Ford and Nelson, 1978) for the calculation of bremsstrahlung spectra and for the generation of secondary electrons in water. For several combinations of target materials and flattening filters, they noticed a linear correlation between the depth of the dose maximum in water and the average energy of the photon spectrum exiting the linac for 10–25 MV photon beams. They concluded that Monte Carlo simulations are a powerful tool for designing target and flattening filter.

Starting in the late 1980s, workers from the National Research Council of Canada (NRCC) in Ottawa devoted several studies to Monte Carlo modeling of bremsstrahlung production in...
linac targets. Bielajew et al. (1989) implemented bremsstrahlung angular sampling from the Koch and Motz distribution (1959) in EGS4. The Koch and Motz (KM) angular distribution was found to lead to a significant difference in the degree of self-absorption in the target compared to using a fixed angle of $m_e c^2/E_0$ (a common approach in many Monte Carlo codes, at one time). This work was complemented by several careful experimental studies (mostly by Faddegon) to determine the energy and angular distribution of emitted photons by targets (Faddegon et al., 1990, 1991). They found reasonable, but not perfect, agreement between simulated and measured spectra. The accuracy of cross sections for bremsstrahlung production remains a study subject today (Poon and Verhaegen, 2005).

5.2.1.1.2 Photon Beam Spot Size in Radiotherapy Accelerators

A point that deserves special attention is the focal spot size of the photon beam or the primary electron beam. The latter is a crucial parameter in Monte Carlo simulations, which influences calculated dose and fluence distributions, as will be demonstrated later. Munro and Rawlinson (1988) were among the first to estimate the size of the x-ray source in linacs. A slit camera, in combination with tomography techniques, was used to project an image of the photons that are emitted under small angles with the central beam axis on a diode detector. Determining the size of the photon source this way is also a direct measure for the spot size of the primary electron beam hitting the target. The presence of a flattening filter does not disturb the spot size measurement since only those photons that go through it without interacting can contribute to the image. For 6–25 MV linacs, they found that the size and shape of the x-ray source can differ from machine to machine, with mostly an elliptical shape with the long axis perpendicular to the gantry axis. The full width half maximum (FWHM) of the measured spots varied between 0.7 and 3.3 mm. Similar studies have been reported (Lutz et al., 1988; Loewenthal et al., 1992; Treurer et al., 1993; Schach von Wittenauf et al., 2002). Jaffray et al. (1993) conducted extensive measurements of focal spot sizes for 6–25 MV photon beams from seven different linacs, using Munro and Rawlinson’s tomographic technique (Munro et al., 1988). They reported measured elliptical spot sizes with an FWHM of 0.5–3.4 mm with eccentricities of 1.0–3.1. They concluded from repeated measurements that the long-term stability of the spot size is mostly determined by the linac design and not so much by linac tuning. They noticed that different energies on the same linac can have different source spot centers: a shift of 0.8 mm was noted for 6 and 18 MV beams on a Varian Clinac 2100C. Focal spots have also been reported to move during the start-up phase of irradiations by up to 0.7 mm in the gun-target direction for Elekta linacs (Sonke et al., 2003). Recently, a method was proposed to estimate the size and the shape of the focal spot from measured dose profile data (Wang and Leszczynski, 2007).

5.2.1.1.3 Photon Target Simulation for Radiotherapy Imaging

A final topic we will discuss briefly in this section is the use of Monte Carlo simulations for the optimization of radiotherapy accelerator targets for portal imaging, which is also relevant for megavolt cone beam CT imaging. Simulation of the imaging detector is of relevance here but detailed discussion of this is not within the scope of this review. Figure 5.3 shows the complete simulation geometry, including the patient and the portal imager downstream. The figure also shows a simulated versus a recorded portal image of a contrast phantom. A linac target for imaging would typically consist of a lower atomic number material than a therapy target. This results in a photon spectrum that is considerably softer, which is needed for imaging purposes due to the imaging contrast at low photon energies.

In some of the older studies (Galbraith, 1989; Mah et al., 1993), detailed experimental work and simulations led to advocating a thin low-Z target (Be or graphite) in a linac with the flattening filter removed to enhance the fraction of photons below 150 keV for the purpose of improving the quality of portal imaging. These studies also demonstrated that a large fraction of the low-energy photons coming of the special target are absorbed in the phantom, thereby reducing the sought improved imaging contrast. Several others (Ostapiak et al., 1998; Tsechanski et al., 1998) used EGS4 simulations of 2–10 MeV electron beam interacting with a range of target materials (Be, C, Al, Ti, Cu). For 4 and 10 MeV electron beams, the optimum Cu target thickness to maximize the integrated photon fluence was found to be 1.5 and 4 mm (Tsechanski et al., 1998), respectively (these targets cannot be considered “thick” as defined above). Somewhat at variance with the previously cited studies (Galbraith, 1989; Ostapiak et al., 1998), Tsechanski et al. (1998) find that there is not much motivation to use imaging targets with an atomic number below Al. This is because in thin targets indeed, more low-energy photons are produced for lower-Z materials but these will mostly be attenuated in the imaged patient and, therefore, contribute to an increased dose. Thin Al or even Cu targets were recommended by these workers.

Building on these works, Flampouri et al. (2002, 2005) presented a comprehensive experimental and simulation study of an optimized portal imaging beam, which was later refined (Roberts et al., 2008). They used the BEAM/EGS4 MC code for detailed simulations of a complete linac treatment head, including a specially designed contrast phantom and several imaging systems. Optical contrast derived from simulations and experiments was found to be in good agreement (Figure 5.3). Recently, Monte Carlo techniques were used for the study of low-atomic-number targets for MV cone beam CT (Connell and Robar, 2010; Robar et al., 2009; Faddegon et al., 2010).

5.2.1.2 Flattening Filter

In this section, we turn our attention to the next component in a linac downstream from the target that has a major influence on the beam: the flattening filter. Flattening filters are designed to generate flat dose distributions at a certain depth in water. We will highlight a number of Monte Carlo simulation studies that provided insight in the influence of the usually complex-shaped flattening filter on photon fluence distributions (see Figure 5.2). It has to be pointed out that there are linacs that do not use
flattening filters. The Microtron MM50 accelerator, tomotherapy machines, and some modern volumetric arc therapy devices do not employ flattening filters (for a recent example of a Monte Carlo study on one of these machines, see (Kry et al. 2010)). McCall et al. (1978) were probably among the first to report the now well-known off-axis differential softening of the photon spectrum, based on Monte Carlo simulations of simplified flattener models made of Al, Ni, and W for a 25 MV photon beam. Since Al was found to cause the largest softening, they recommended that flattening filters for high-energy photon beams should be made of medium-\(Z\) materials such as Cu or steel. Mohan et al. (1985) simulated 4–24 MV photon beams from a number of different linacs. They were the first to build accurate models of the flattening filters (presumably made of steel) by developing a special geometry package that allowed easy modeling of the complex geometry (another forerunner of BEAM). They demonstrated that flattening filters cause significant spectral hardening both on and off the beam axis. For example, a 15 MV beam with no flattening filter present yielded average photon energies at 100 cm from the source of 2.8 and 2.5 MeV, respectively, at the central axis and in an annular scoring region between 10 and 25 cm. These values changed to 4.1 and 3.3 MeV when a flattener was added, increasing the differential off-axis spectral softening significantly. It can be seen from their work that central axis and off-axis spectra differ mostly for energies below 1 MeV. They also reported that the average photon energy from a linac is lower than the one-third of the nominal energy as then was—and today still is—commonly assumed.

A generic model of a linac (“McRad”) based on the EGS4 code was presented by Lovelock et al. (1995) at around the same time BEAM was released. For a 6 MV beam, these authors observed a 10% off-axis softening at 100 cm for an unflattened beam whereas adding a flattening filter augmented this to about 30%. The fact that all these studies report a relatively small off-axis softening for unflattened beams is related to the observation that the bremsstrahlung spectrum from a thick target is fairly isotropic within the angular range of the photons that can reach the patient, whereas the bremsstrahlung intensity is highly anisotropic, as discussed before (Section 5.2.1.1.1). Faddegon et al. (1999) used BEAM simulations to successfully redesign a flattening filter of a clinical accelerator to obtain larger flat 6 MV fields. This involved changing the material of the flattener from steel to brass.

FIGURE 5.3  Schematic representation of a linac simulation, including a patient and a portal imager below the patient. The lower-energy photon spectrum (dotted line) used for imaging purposes is compared to the therapy photon spectrum (full line) in the graph. The bottom panels compare a recorded portal image (left) of a contrast phantom to a simulated one (right). (Work by S. Flampouri, University of Florida Proton Therapy Institute.)
5.2.1.3 Monitor Ion Chamber Backscatter

In some clinical linacs, the signal from the beam monitor ion chamber is affected by the position of the secondary movable beam collimators (jaws). This only occurs in linacs where the distal monitor chamber window is sufficiently thin, where no backscatter plate is present and where the distance between chamber and the upper surface of the collimators is small enough. In that case, particles backscattering from the movable collimators can deposit charge in the monitor chamber, in addition to particles moving in the forward direction in the beam. In a small field, more backscatter from the collimators will occur than in a large field. This means that the monitor chamber will reach its preset number of monitor units (MU) quicker and terminate the beam. This will cause the linac output to decrease with decreasing field size. The magnitude of this effect is usually limited to a few percent but for some linac types, larger effects have been reported. The decreased output is automatically included in output factor measurements, but when Monte Carlo simulations are used to determine output factors, the effect has to be taken into account separately. This applies to all cases where outputs from fields of different sizes are combined, so it also plays a role in modeling dynamic (or virtual) wedges and possibly even in intensity-modulated radiation therapy (IMRT) modeling. This only is possible in certain types of linacs, but one should be aware of this effect that may cause simulations to disagree with measurements.

A few studies have used Monte Carlo techniques to investigate the backscatter effect. Liu et al. (1997a, 2000) performed several detailed simulations of Varian linacs to study the differential effects of the different collimators on monitor backscatter. They concluded that the width of the jaw opening is important to determine the backscatter correction, whereas the actual off-axis position of the field is not important. Verhaegen et al. (2000) modeled Varian 6 and 10 MV photon beams, including a detailed model of the monitor ion chamber. By tagging particles and selectively transporting photons and electrons, it was found that electrons cause most of the backscatter effect. A spectral analysis of the forward and backward moving particles was shown. In this study it was also found that in this linac the backscatter from MLC is negligible at the level of the monitor chamber. In the two studies cited where MC simulations and measurements were compared (Liu et al., 2000; Verhaegen et al., 2000), good agreement was observed.

5.2.1.4 Wedges

Inserting a physical wedge in a photon beam alters the beam characteristics significantly. Not only is the dose distribution modified by attenuation and scatter from the wedge, but the photon spectrum is also affected. Usually, spectral hardening is seen below some parts of the wedge. When a moving linac jaw is used to form a dynamic (or virtual) wedge, the photon spectrum is much more similar to the open-field spectrum. For very-high-energy beams (>20 MV), beam wedges may cause beam softening due to, for example, pair production and annihilation.

Monte Carlo simulations can be used to model physical or dynamic wedges. When modeling a physical wedge, great care has to be taken that the model for the wedge is exact. Not only the exact shape of the wedge has to be implemented, but the composition and density also have to be known exactly. In particular, steel wedges have been known to be difficult to model because of uncertainties in the manufacturer's specified composition.

Liu et al. (1997b) extended their dual-source model to include a wedge model for a 6–18 MV photon beams of a Varian 2100C linac. Steel and lead wedges with angles from 15° to 45° were modeled. The wedge was seen as part of the patient-dose calculation. Special photon dose kernels to take into account the effect of the wedge were added in their convolution/superposition dose calculation method. These kernels were calculated using MC techniques in bimaterial spheres (lead and water). The beam hardening by the wedge was included in the model. Their dose calculations were in agreement with full Monte Carlo simulations, except in the build-up region because of lack of electrons coming from the wedge in their source model. It was found that the wedge-generated photon fluence contributed significantly to the total dose: for example, a 45° wedge in a 20 × 20 cm² 10 MV photon beam caused 8.5% of the total dose at the central axis at a depth of 5 cm. By distinguishing between annihilation, bremsstrahlung, and Compton-scattered photons from the wedge, it was found that the latter was responsible for the majority of the wedge-generated dose. The wedge was found to produce a near-Gaussian-shaped lateral dose profile.

Schach von Wittenau et al. (2000) used crude Monte Carlo wedge models and found that wedge-generated bremsstrahlung photons carried about 20% of the outgoing energy in a 10 MV beam. Li et al. (2000) presented their Monte Carlo code MCD OSE, which takes wedges into account as a part of the patient-dose calculation. Spezi et al. (2001) modified DOSXYZ (now updated to DOSXYZnrc (Walters et al., 2009)) to include a rectilinear voxel geometry module and the ability to collect phase-space information behind the wedge. They built a model for the lead wedges of their 6 MV photon beam of a Varian 2100CD linac. A step-resolution of 1 mm in the wedged and beam direction was found to be sufficient for the dose calculations.

Van der Zee and Welleweerd (2002) introduced a new component module in the BEAM code that allows modeling complex wedges such as the Elekta internal wedge. Figure 5.4 shows the CM WEDGE that was used by the authors to model the internal wedge in their 6–10 MV photon beams of an Elekta SL linac. They found that the presence of the wedge altered the primary and scattered photon components from the linac significantly: beam hardening by 0.3 and 0.7 MeV was observed for the two components, respectively. They also noted that the wedge-generated photons were mostly generated close to the distal edge of the wedge. Figure 5.4 also shows the photon energy fluence originating from various sources in the linac, including scattered photons from the wedge. Monte Carlo techniques are unique in extracting such information.

In building Monte Carlo models for dynamic wedges, the time-dependent movement of the jaw has to be included. The
studies we will discuss here laid the foundations for modeling motion of MLC in IMRT, to be discussed in the next section. Verhaegen and Das (1999) built a BEAM MC model for the Virtual Wedge of 6–10 MV photon beams of a Siemens MD2 linac. The wedge was modeled by a discrete sum of a large number of open fields. By comparing calculated photon spectra from the heel, toe, and central regions, they found that the photon spectrum of the Virtual Wedge was only slightly hardened, with no significant difference for the three regions. In contrast, 60° physical wedges of tungsten introduced a significant beam hardening for the 6 MV beam across the whole wedge. Both the transmitted and wedge-generated photons were found to be harder for a tungsten wedge than for an open field. Steel wedges were found not to alter the spectrum significantly.

A full dynamic simulation technique to model a Varian enhanced dynamic wedge was introduced by Verhaegen and Liu (2001). This is the first report of a fully dynamic Monte Carlo simulation of a linac component. They used the BEAM Monte Carlo code to model 6–10 MV dynamically wedged photon beams of a 2100C Varian linac by sampling the position of the moving jaw from the segmented treatment table. More information is given in Chapter 7. They termed their technique the “position-probability-sampling method,” which is implemented in a BEAM CM, DMLCQ. The varying backscatter to the monitor chamber during the jaw movement was included in the model. The authors obtained excellent agreement between measured and calculated wedged dose profiles. Shih et al. (2001) presented detailed information on photon and electron spectra and angular distributions produced by dynamic and physical wedges.

Monte Carlo models for other beam modifiers such as blocks, compensators, and so on were discussed in the literature (Jiang and Ayyangar, 1998; Li et al., 2000; Spezi et al., 2001; Ma et al., 2002).

5.2.1.5 Multileaf Collimators and Intensity-Modulated Radiation Therapy

MLC, consisting of banks of tungsten leaves that allow shaping of conformal or intensity-modulated fields in IMRT, are among the most challenging geometrical structures in a linac to model. The MLC leaves can have very complex designs for the leaf ends and the leaf edges. An additional complication is that MLC leaves can move during beam delivery in step-and-shoot and dynamic treatments. The only way that these complex beam-shaping devices can be fully taken into account is by Monte Carlo simulation, but this is a far from trivial task. Furthermore, significant differences in radiation spectrum in
different sections of IMRT fields could have significant effects in radiobiology or for film measurements often used in IMRT verification (Liu and Verhaegen, 2002).

The Monte Carlo research for modeling MLCs has mainly been focused on two aspects: exactness of the geometrical models and methods to improve calculation efficiency. Because different linac manufacturers have implemented different designs of MLCs, there is no generic MLC model available in codes like BEAM. There are also differences in how IMRT beams are delivered in different linacs. All of this has to be taken into account in a Monte Carlo model for IMRT. We would also like to point out that the mass density of the tungsten used in the MLC leaves usually has to be ascertained for every linac, due to significant variations in different linacs. Backscatter from the MLC leaves to the monitor chamber does not play a significant role due to the large distance of the MLC to the chamber (Verhaegen et al., 2000), the presence of a backscatter plate (Hounsell, 1998), or the thickness of the monitor chamber downstream window thickness (Verhaegen and Das, 1999).

An early Monte Carlo study of MLC leaf geometries was Küster (1999). She used the GEANT3 Monte Carlo code to investigate the characteristics of two experimental designs of MLCs for a Siemens Primus linac with the purpose of determining the leaf end shape required to optimize dose penumbras. In this feasibility study, leaf leakage was also investigated but good agreement between calculations and measurements was not obtained.

One of the first to show good agreement between measured and simulated MLC dose distributions was De Vlamynck et al. (1999). They built a model for a 6 MV MLC–shaped photon beam from an Elekta SL25 linac. The rounded shape of the MLC leaves was approximated by using a stack of MLC modules in BEAM simulations. Interleaf leakage was not modeled in this study. Figure 5.5 gives an example of the good agreement obtained between calculated and measured dose profiles and output factors for small fields with an off-axis position of up to 12.5 cm. No significant spectral differences were noted when large and small fields were compared. In the same paper, a new component module for the BEAM code, MLCQ, was introduced. This allowed modeling of MLC leaves with curved leaf ends.

Ma et al. (2000) modified BEAM/DOSXYZ to allow simulation of multiple-beam fixed-gantry IMRT treatments. The information in the linac MLC leaf sequence file was used to derive a fluence map from which particles were sampled during the simulations. The statistical weight of the particles was altered using the information in the fluence map. The MLC leakage was incorporated by reducing the particle statistical weight for the MLC blocked sections. Their model did not include the effects of the leaf shape, tongue-and-groove design, and scattering in the MLC material, but still they reached an agreement within 2% for measured and calculated dose distributions. For complex IMRT plans, they showed differences between their MC system and a conventional pencil-beam algorithm of up to 20%. The same group of workers (Deng et al., 2001) later studied the MLC tongue-and-groove effect on IMRT dose distributions. They used a ray-tracing approach to modify the fluence maps so that the presence of the tongue-and-groove geometry was accounted for. For single-field IMRT dose differences due to the tongue-and-groove effect of 4.5% were noted; when more than five fields were used the differences fell below 1.6%. They remark that the effect of the tongue-and-groove geometry is probably insignificant in IMRT, especially when organ/patient movement is considered. The same workers (Kapur et al., 2000) introduced a new BEAM component module, VARMLC, that allows modeling the tongue-and-groove effect of some MLCs.

Work by another group (Keall et al., 2001; Kim et al., 2001; Siebers et al., 2002) introduced an elegant model to take the MLC geometry into account in Monte Carlo simulations without actually having to model any particle transport in the MLC material. In Keall et al. (2001) the MLC was compressed in a thin layer in the beam direction. Geometrical paths of photons passing through the MLC were determined for particles sampled

![FIGURE 5.5](Image)

Comparison between measured and Monte Carlo calculated MLC dose profiles (10 × 1 cm² field, 12.5 cm off-axis, left panel) and output factors (10 × 1 cm² field at various off-axis distances, right panel). The agreement is good in general, especially for relative doses but it also shows that absolute dose output for these small fields is hard to calculate correctly. (From De Vlamynck, K. et al. 1999. *Med Phys* 26:1874–1882. With permission.)
from a phase-space file above the MLC. The statistical weights of the particles are modified according to the probability that the particles reach the bottom of the MLC geometry, where a new phase-space file is constructed. The model took into account intraleaf thickness variations and the curved shape of the leaf end but omitted interleaf leakage and charged particle transport; hence, production of bremsstrahlung photons and annihilation photons. First-scattered Compton photons were included but transport across leaves was not considered. This model was extended (Siebers et al., 2002) to take the leaf edge effect into account. Cross-leaf photon transport and charged particle transport in the leaves was still ignored. They make the interesting observation that it may suffice to model only first Compton scatter and omit electron transport in other linac components upstream from the MLC. The same group also fully simulated MLC geometries (Kim et al., 2001). The leaf ends, leaf edges, mounting slots, and holes were included in the study. They investigated radiation leaking through the MLC. Depending on the field size defined by the jaws above the MLC and the photon energy they found that fully blocked MLC fields had a radiation leakage dose of about 1.5–2%, compared to the dose in an open field. It was observed that the MLC generated photons add a broad background dose to the transmitted radiation (Figure 5.6). Electrons emitted from the MLC cause up to 35% of the surface dose (Kim et al., 2001) in an 18 MV MLC blocked photon beam. For 6 MV photons, significant hardening of the photon spectrum behind the MLC was observed, resulting in a significant shift of the depth dose curve. No such effect was noted for 18 MV photons.

Recently, a method to greatly speed up MLC simulations was reported (Brualla et al., 2009) in which sophisticated variance reduction techniques are used to automatically determine geometrical regions that require detailed transport and others where transport may be performed in an approximate fashion. They also show that modeling only single Compton scatter in the MLCs does not suffice.

Liu et al. (2001) were the first to implement fully dynamic MLC motion to allow modeling step-and-shoot or dynamic IMRT delivery. This resulted in a new BEAM component module, DMLCQ. During sampling of particles for transport, the leaf positions were randomized according to the number of MU in each field segment. This information was derived from the linac leaf sequence file. A correction for the difference between nominal and actual leaf position was required. Figure 5.7 compares Monte Carlo calculations and measurements for an IMRT field delivered in step-and-shoot mode. One of the most detailed MLC models ever published (Heath and Seuntjens, 2003), based on Liu et al. (2001), included leaf leakage, mounting screws, leaf divergence, support railing groove, leaf tips, and dynamic motion.

Li et al. (2001) built a model for intensity-modulated arc therapy (IMAT) delivery, which involves an Elekta linac SL20 arcing around the patient while the MLCs are moving. An additional complication is that the SL20 uses its internal wedge for some fractions of the delivery. They obtained good agreement with measurements but found up to 10% difference between Monte Carlo and conventional dose calculation techniques for clinical IMAT fields. Many more papers on Monte Carlo simulations for arc therapy have been published since (Bush et al., 2008; Teke et al., 2009); more on this topic will be discussed in Chapter 7.

The effect of the MLC on portal dose distributions was studied using Monte Carlo techniques (Fix et al., 2001; Spezi and Lewis, 2002).

### 5.2.2 Full Linac Modeling

In the following two sections, we will describe complete linac models for photon beams. The photon beam models can be basically divided into two categories. The first approach is to do complete simulations of linacs and either use the obtained particles directly in calculations of dose, fluence, and so on in phantoms, or store phase-space information about these particles at possibly several planar levels in the linac for further use. The phase-space can be sampled for further particle transport in the rest of the geometry (linac or phantom). If the phase-space is situated at the bottom of the linac, this then effectively replaces the linac and becomes the virtual linac. We will call this the phase-space approach. In Figure 5.1, two levels where phase-space information might be collected are indicated (levels AA' and BB'). A particularly elegant approach couples BEAMnrc directly to DOSXYZnrc simulations, thereby eliminating the need to store intermediate phase-space files.

The second approach calculates particle distributions differential in energy, position, or angle in any combination of these in one-, two-, or three-dimensional histograms. This is usually done for several subsources of particles in the linac and is termed a Monte Carlo source model. The linac is then replaced by these subsources that constitute the virtual linac. The source model approach often starts from information
collected in phase-spaces in the linac. Source models inevitably have approximations since they condense individual particle information in histograms, whereas all the information on individual particles is preserved in the phase-space approach. The correlation between angle, energy, and position of particles might be partially or completely lost depending on the degree of complexity of the source model. The disadvantage of working with phase-spaces is the large amount of information to be stored and the slower sampling speed during retrieval of all this information. Also, approximations in the linac model will result in uncertainties in the phase-space approach. Source models often involve data smoothing, so they usually result in less statistical noise.

5.2.2.1 Complete Linac Simulations

As mentioned before, Patau et al. (1978) pioneered the Monte Carlo simulation of a more or less complete model of a photon linac. They presented energy and angular distributions of photons before and after a flattening filter but attempted no further analysis of the influence of the linac components on the photon beam. In another early study (Nilsson and Brahme, 1981), linac components were sequentially made transparent (i.e., no interactions were allowed in them) to study the influence of the components on the particle beam. They found that in a 21 MV beam, the thicker flattening filter, combined with the more forward-peaked bremsstrahlung angular distribution, caused most of the photon scatter. In a 6 MV beam, the primary collimator contributed most to the scatter, corresponding to the larger angular bremsstrahlung distribution for lower primary electron energies.

Mohan et al. (1985) presented a much-quoted EGS3 simulation study of a set of Varian linacs producing 4–24 MV photon beams. The model included a target/backing, a primary collimator, a flattening filter, and a secondary collimator system. They noted that off-axis photon spectra were softer than at the central axis. The collimating jaws had no significant effect on the energy and angular photon distributions. The beam was mostly determined by the target and the flattening filter. For 15 MV photons, they found that when photons reached a plane at 100 cm from the source, 93.5% had never scattered whereas 2.8%, 3.5%, and 0.2% suffered scattering interactions in the primary collimator, flattening filter, and secondary collimator system, respectively. The photon spectra presented in this study are a reasonable approximation; they are therefore still frequently used today, for example, for patient dose calculations. Their angular distributions, however, are only crude approximations.

A model of a 6 MV Siemens linac photon beam including the exit window, target, primary collimator, flattening filter, monitor chamber, and collimating jaws was published by Chaney et al. (1994). They focused on head scatter contributions. Despite the model being very approximate, the calculated scatter contributions are nevertheless found to be in fair agreement with more detailed studies. They found that about 9% of the photons in a large field were derived from scatter. Figure 5.8 shows their scatter-plot of the sites of origin of scattered particles. From this
figure, it is evident that Monte Carlo simulations—besides producing aesthetically pleasing pictures—can provide clear insight into which parts of the linac play an important role as a source of scatter or particle generation.

Libby et al. (1999) give detailed information on how to validate Monte Carlo linac models. As an example of verification of linac geometry, they show how the complex shape of the flattening filter can be reproduced and compared to manufacturer’s data by bombarding it with monoenergetic, monodirectional photons and calculating the attenuation profile. Another approach (Heath and Seuntjens, 2003) for verifying complex geometries is ray-tracing photon tracks in such a way that photon coordinates are stored every time a photon crosses a boundary between MLC leaf material and air. Figure 5.9 gives an example of the geometry that can be visualized in this fashion.

Detailed models of photon linacs were presented by many others (Mazurier et al., 1999; van der Zee and Welleweerd, 1999; Lin et al., 2001; Ding, 2002a; Reynaert et al., 2005; Titt et al., 2006; Vasiliev et al., 2006; Mesbah et al., 2007, 2005; Bednarz and Xu, 2009). Highly detailed studies on the influence of linac components and the characteristics of the primary electron beam were done by Sheikh-Bagheri and Rogers, (2002a,b) and Sheikh-Bagheri et al. (2000). They advocated using in-air dose profiles measured with an ion chamber and a build-up cap to derive the energy and spatial characteristics of the primary electron beam. Figure 5.10 illustrates the sensitivity of the off-axis factor (defined as the ratio of measured dose to water with an ion chamber in air at an off-axis position and at the central axis) to the primary electron energy and the width of the radial intensity distribution of the electron beam. By avoiding phantom scatter as a confounding factor, the authors clearly demonstrate the differential effects of the components in their model. Interestingly, the insensitivity of dose calculations to the energy distribution of the primary electron beam for photon beam simulations is in contrast to electron beams where both build-up and build-down regions are highly sensitive to the primary electron energy distribution.

A general finding is that the average photon energy is below one-third of the nominal energy at the central axis (more so for the highest-energy beam) and that the average energy 20 cm off-axis is decreased by 0.5–2.0 MeV (the greatest decrease usually corresponds to the highest nominal energy). Ding (2002a) noted that the mean energy of the charged particles in a photon beam varies less with off-axis distance than the photon mean energy. A procedure to tune a Monte Carlo linac model was outlined by Verhaegen and Seuntjens (2003).

### 5.2.2.2 Source Models

Early Monte Carlo-based source models were first introduced for electron beams and then later for photon beams (Liu et al., 1997a,b; Ma, 1998). Liu et al. (1997a,b) introduced a dual-photon source model: one source for the primary photons coming directly from the target and a second extra-focal source to describe scatter from mainly the primary collimator and the flattening filter. A third source took the electron contamination into account. Derivation of the source model started by performing a full linac Monte Carlo simulation and phase-space analysis. The obtained distribution of points of origin of photons for a 10 MV beam is illustrated in Figure 5.11. Obviously, it is very important in such a model that the geometrical position of the scatter sources is estimated correctly. Inevitably, replacing the real three-dimensional source distributions with two-dimensional ones is an approximation. The primary photon source at the target was modeled with a radius of 0.1 cm. The effect of the jaws was modeled as eclipsing collimators, thereby ignoring the small amount of scattered radiation the jaws contribute. Backscatter...
toward the monitor chamber was added explicitly, which is often omitted in later works. Despite the approximations in the model (Liu et al., 1997a, b), good agreement was reached between calculated and measured dose distributions and output factors.

Other complex source models can be found in the literature (Schach von Wittenau et al., 1999; Deng et al., 2000; Fix et al., 2000, 2001; Hartmann Siantar et al., 2001; Sikora et al., 2007; Tillikainen et al., 2007). The PEREGRINE group (Schach von Wittenau et al., 1999) performed a very detailed study to investigate the influence of linac components on photon fluence, energy spectra, angular distributions, and so on. They introduced the concept of correlated histograms, which contains information extracted from phase-space files in such a way that correlations between energy, angles, and so on are preserved to some degree. Target photons were found to originate from a small disk source with a radial fluence distribution, the primary collimator was found to generate photons in a ring source close to its upper edge and the flattening filter was found to produce photons uniformly throughout the filter geometry. Figure 5.12 shows the correlated histogram used for the two subsources for a 6 MV beam: target and scattered photons. The authors reached very good agreement between full phase-space-based simulations and their model. A subsource was added to model the electron contamination (Hartmann Siantar et al., 2001) in the PEREGRINE photon Monte Carlo treatment planning system. The jaws were modeled as a masking collimator, which causes the dose outside the penumbra to be up to 15% too low. Backscatter to the monitor chamber was included as an empirical correction.

One of the most complex source models published to date was presented by Fix et al. (2001). Using GEANT3 simulations, they derived a model for a 6 MV linac beam with 12 subsources: one each for the target, flattening filter, primary collimator, and mirror, and two each for all four of the secondary collimators. The two subsources for each secondary collimator represent radiation emanating from the inner side and the lower side of the collimators. It was found that the inner-side subsource generates most particles close to a 1 cm
distance from the top of the collimator, whereas the bottom-side subsource generated most particles close to the field edge. The total contribution from the eight collimator sources was very small and is usually ignored in most linac source models. Data was generated for eight field sizes, which allowed interpolation for arbitrary fields.

The source model approach has to be used with care due to its inherent approximations and potential for systematic error. Its strength lies in the fact that the requirements on data storage are greatly reduced with respect to a full phase-space approach; figures of 400–10,000 have been quoted (Chetty et al., 2000; Deng et al., 2000; Fix et al., 2000). Because of the reduced phase-space reading/writing and because source models inherently involve some degree of fluence smoothing, the calculation time required to achieve a certain specified statistical variation in a dose calculation is also reduced. Monte Carlo-based treatment planning systems often resort to virtual source models.

5.2.2.3 Absolute Dose Calculations (Monitor Unit Calculations)

Dose calculations based on Monte Carlo simulations of linacs (or any other device) usually result in doses expressed in Gy/particle. In the BEAMnrc package, the term “particle” refers to the initially simulated particles, so in a linac simulation, the calculated doses are in Gy per initial primary electron hitting the target. Even in simulations that are done in several parts, involving multiple phase-space files at different levels in the linac, this information is preserved in the phase-space files. This ensures correct normalization of the doses per initial particle. Other Monte Carlo codes use different schemes or leave this up to the user. This absolute dose can be related easily to the absolute dose in the real world. By running a simulation that exactly matches an experimental setup to determine absolute dose, one can obtain the ratio of the measured and calculated doses, which are expressed in units of [Gy/MU]/[Gy/particle], where MU stands for monitor unit. The latter corresponds to the reading of the monitor ionization chamber in the linac. This ratio may serve as conversion factor for all beams produced by the linac for that particular photon energy. By obtaining these conversion factors for all linac energies, Monte Carlo calculated doses may be converted to absolute doses as used in radiotherapy practice. This approach is also valid for electron beams (see Chapter 6), and is employed by most users. The reference setup that is frequently modeled is a dose determination in an open 10 \times 10 \text{cm}^2
field in water, but in principle any setup, not necessarily using open fields, can be modeled to determine the conversion factors for each photon energy. This approach was described in more detail in the literature (Leal et al., 2003; Ma et al., 2004) with special attention to IMRT treatments. A different approach that employs the relation between the simulated monitor chamber reading, the initial number of particles impinging on the target, and the field size was introduced by Popescu et al. (2005). This elegant method requires attention to the issue of backscatter of particles to the monitor chamber (Section 5.2.1.3). It does not rely on renormalization to measurements and relative output factors. Although in principle a superior approach, it is not often used in practice, as far as we know.

5.2.3 Stereotactic Beams

Linac-based stereotactic irradiation of intracranial and extracranial lesions is a frequently employed radiotherapy technique. Monte Carlo simulation of the collimated, narrow photon beams (10–30 mm at the isocenter) is therefore important in two areas. First, knowledge of detailed beam data allows a proper understanding of detector response that is often affected by the absence of lateral electronic equilibrium. Second, treatment planning systems may employ Monte Carlo techniques to reliably take into account tissue heterogeneities near small targets. This implies that the basic properties of the beam should be accurately calculated using geometric treatment head information.

Several Monte Carlo studies on stereotactic beams (Kubsad et al., 1990; Sixel and Podgorsak, 1993, 1994; Sixel and Faddegon, 1995; Verhaegen et al., 1998; Chaves et al., 2004; Belec et al., 2005; Jones and Das, 2005; Ding et al., 2006; Scott et al., 2008, 2009; García-Pareja et al., 2010) and a recent review paper (Taylor et al., 2010) have been published. Studies have focused on the accuracy of dose calculations in small fields, photon scatter, photon and electron spectra, and correction factors for measurements. A full simulation (Verhaegen et al., 1998) of a dedicated 6 MV stereotactic linac with field sizes ranging from 5 to 50 mm showed that photon spectra are influenced by the small beams; mean photon energies at $d_{max}$ in water were found to vary between 2.05 MeV for a 5 mm cone to 1.65 MeV for a 5 cm cone. They also observed that Spencer–Attix stopping-power ratios at the central axis were depth independent, which is an important information for measurements. Another study simulated a small-field dynamic radiosurgery unit capable of producing 1.5 mm pencil beams (Paskalev et al., 2002, 2003) and focused on aspects of lateral electronic disequilibrium. Differences of up to 60% were found between Spencer–Attix stopping-power ratios and dose-to-water and dose-to-cavity ratios; a correction procedure was proposed. It has been demonstrated (Scott et al., 2008, 2009) that small-field output factors are very sensitive to the dimensions of the primary electron beam, which suggests a way to commission this part of the linac geometry.

Some have suggested that Monte Carlo dose calculations for treatment planning in stereotactic beams may not be needed (Ayyangar and Jiang, 1998), while others (Solberg et al., 1995) have argued that Monte Carlo simulation is required for reliable small-field dose calculation in regions of tissue heterogeneity for which, in conventional planning systems, often only the primary photon beam component (and not the scatter component) is considered. Full Monte Carlo simulations are preferred for small fields but virtual source models have also been proposed (Chaves et al., 2004).

As an example, Figure 5.13 shows the level of detail that is needed to model a linac with a micro-MLC for stereotactic irradiation. The geometric accuracy of the leaf ends and sides, and the composition of the leaves are very important. The figure also shows the level of agreement with measurements that can be achieved by an accurate Monte Carlo model. Conventional dose calculation was shown to be considerably less accurate.

Monte Carlo studies of a CyberKnife robot-mounted small-field linac have also been published (Yamamoto et al., 2002; Araki, 2006; Sharma et al., 2010; Wilcox et al., 2010). A recent study introduced an exotic technique to enhance the simulation efficiency of small fields. The “ant colony” method (García-Pareja et al., 2010) was employed whereby geometric cells are overlaid with an importance map that is derived in an analogous fashion in which ants leave pheromone trails to food sources. In their approach, photons that reach regions of interest determine the importance maps so that subsequent photons may be simulated more efficiently.

5.2.4 Contaminant Particles

5.2.4.1 Electrons

Electron contamination in photon beams has been studied by numerous authors (Padikal and Dey, 1978; Nilsson and Brahme, 1979; Ling et al., 1982; Mackie and Scrimger, 1982; Petti et al., 1983a; Rustgi et al., 1983; Yorke et al., 1985; Gerbi and Khan, 1990; Lamb and Blake, 1998; Zhu and Palta, 1998; Malataras et al., 2001; Ding, 2002b; Ding et al., 2002; Yang et al., 2004; Abdel-Rahman et al., 2005; Lopez Medina et al., 2005; Mesbah, 2009; Sikora and Alber, 2009; Ververs et al., 2009). Electrons in the treatment beam are generated either in the high-$Z$ components of the linac or in the air. It is known that electrons in the photon beam are responsible for a significant fraction of the surface dose in a phantom or in patient. The dose from contaminant electrons drops off rapidly with depth and is usually insignificant at the depth of maximum dose of the photon beam, $d_{max}$ (for an overview of experimental findings, see Verhaegen and Seuntjens, 2003).

Petti et al. (1983a,b) were the first to dedicate a Monte Carlo study to electron contamination in photon beams. They observed that 70% of the electrons at the phantom surface were generated in the flattening filter and monitor ion chamber, 13% from the collimating jaws, and 17% in the air for a source to surface distance (SSD) of 80 cm. For an SSD of 400 cm, these fractions changed to 34%, 5%, and 61%, respectively. They concluded from this that at all SSDs the contribution from the flattening filter and monitor chamber is significant whereas the contribution from air-generated electrons is only important at very large SSD. The former was also found to cause the shift of $d_{max}$ with field size because the collimating jaws act as an eclipsing absorber. At an SSD of 80 cm, these electrons had a most probable energy of...
1.5 MeV (with 40% of them having an energy exceeding 5 MeV) and a forward-peaked angular distribution with 60% of the electrons moving at an angle with respect to the beam axis of less than 16°. These findings explain why these electrons still play a significant role at very large SSDs. The electrons generated in the collimating jaws, which are commonly assumed to be important, were found to play a minor role.

Some Monte Carlo studies (Fix et al., 2001; Malataras et al., 2001) showed energy spectra, energy fluence distributions, and mean energy distributions for the electrons reaching the phantom or bottom of the linac. They found that the energy spectra of electrons generated in the flattening filter and in air are very different: the former has a broad energy spectrum with an average electron energy comparable to the average photon energy. The latter has a narrow energy distribution, peaked at low energies. These results may be very linac dependent. It was pointed out (Malataras et al., 2001) that the increasing probability of pair production causes the flattening filter to become the dominant electron source at high photon energies (see Figure 5.14).

For quite a long time, Monte Carlo studies of surface dose in photon beams showed significant differences with measurements (van der Zee and Welleweerd, 1999; Hartmann Siantar et al., 2001; Ding, 2002b; Sheikh-Bagheri and Rogers, 2002a,b). Only with very detailed linac models and taking into account that radiation detectors may need various response corrections close to an air/phantom interface, a few workers (Sheikh-Bagheri et al., 2000; Abdel-Rahman et al., 2005) succeeded in achieving good agreement. These studies showed that Monte Carlo dose calculations near surfaces are often more reliable than measurements, especially if corrections are not taken into account.

5.2.4.2 Neutrons
It is known that photons with an energy exceeding about 8 MeV can produce neutrons in the target and collimating structures. For an overview of experimental techniques, we refer to AAPM Report 19 (1986). Several authors used Monte Carlo techniques to simulate 15–50 MV photon beams from a variety of linacs (Ing et al., 1982; Agosteo et al., 1993; Gudowska et al., 1999; Ongaro et al., 2000). They included methods to estimate ($\gamma, n$) production in linac components (mostly the target) and phantom geometries. Monte Carlo codes that can handle photo-neutron production are, for example, MCNP(X). It has been determined

![Figure 5.13](image-url)
Radiotherapy Kilovolt X-Ray Units

Kilovolt x-ray tubes have been used in radiotherapy for many decades but are now being used less frequently for cancer treatment. Linac electron beams and brachytherapy sources have largely replaced the use of kV x-rays for superficial or skin lesions, respectively. This by no means implies that there is no role for Monte Carlo modeling of kV x-ray tubes. In diagnostic imaging, Monte Carlo methods are commonly used to model x-ray tubes for planar or CT imaging. A recent example of CT scanner modeling versus Compton scatter spectroscopy is Bazalova and Verhaegen (2007). The literature contains many examples of Monte Carlo modeling of x-ray tubes for diagnostic and mammography applications (Boone et al., 2000; Verhaegen and Castellano, 2002; Ay et al., 2004) for example, dosimetric purposes. Detailed discussion of these mostly diagnostic applications is outside the scope of this chapter. There are some specialized applications of kV x-rays in radiotherapy that are under investigation and where Monte Carlo techniques play an important role, such as contrast-enhanced radiotherapy, in which kV x-rays are used to bombard targets laden with dose enhancers such as high atomic number contrast media (Verhaegen et al., 2005) or gold nanoparticles (Jones et al., 2010). To model the interaction of kV x-rays with high atomic number materials accurately, a number of physical processes must be implemented, which makes photon transport in the kV energy range harder than in the MV energy range where Compton scatter is the dominant interaction. Examples of things that need to be implemented in the transport models and which largely can be ignored at MV energies are electron impact ionization, Rayleigh scattering, photo-electron angular distributions, and fluorescent decay. Another specialized area is the modeling of miniature x-ray tubes that are small enough to be implanted in lesions to irradiate them from within. These devices often consist of very thin transmission x-ray targets made of a composite of various high-atomic-number materials, stressing the importance of the accuracy of bremsstrahlung cross sections and the modeling of electron impact ionization. Examples are the work of Yanch and Harte (1996) on brain applications and Liu et al. (2008) on electronic brachytherapy. Another special application is small animal irradiation for preclinical studies that is usually done with kV x-rays to avoid the build-up regions of MV photon beams in small animals such as mice (Tryggestad et al., 2009). Monte Carlo modeling also plays a role in beam simulation and dose calculation. Since these applications are outside the field of external beam radiotherapy, they will not be discussed in detail here.

X-ray spectra are used in many different applications, for example, in studies in radioprotection, radiobiology, radiation quality, and response of radiation detectors. Therefore, several authors created methods to generate x-ray spectra as a function of target angle (most efforts focused on tungsten targets), accelerating voltage and filtration. Whereas earlier work (Birch and Marshall, 1979; Iles, 1987) employed analytical techniques, the most recent work relies on Monte Carlo methods for detailed electron penetration in the target (Poludniowski, 2007; Poludniowski and Evans, 2007). The SpekCalc code (Poludniowski et al., 2009) uses a graphical user interface to perform a very fast, yet accurate, x-ray spectrum calculation for tungsten targets bombarded by electrons in the range 50–150 keV.
In the late 1990s, the BEAM code was used to simulate entire radiotherapy x-ray tubes (50–300 kV), including effects such as tube voltage ripple (Verhaegen et al., 1999). The results were compared to spectral measurements and another Monte Carlo code, MCNP. Both codes had problems with the characteristic x-ray lines due to limitations in the modeling of electron impact ionization. Compared to the measurements, there was a clear overestimation of the low-energy component in the photon spectrum at the expense of contributions in the high-energy part of the spectrum. Other workers modeled x-ray tubes for radiotherapy or diagnostic applications (Omrane et al., 2003; Ay et al., 2004; Ay and Zaidi, 2005; Kakonyi et al., 2009; Taleei and Shahriari, 2009) and special attention was paid to variance reduction techniques (Mainegra-Hing and Kawrakow, 2006). A recent study modeled in great detail the backscatter from electrons at x-ray targets, and the subsequent reentry of the electrons due to interaction with the electric field inside the x-ray tube (Ali and Rogers, 2008). This effect causes the well-known extrafocal radiation of which the authors demonstrate that it may influence the spectral shape (therefore, the half-value layer) and the air kerma. Figure 5.15 shows the magnitude of the backscatter coefficients for first- and higher-generation electrons obtained from Monte Carlo simulation. For applications with a hard filtration and tight beam collimation, the effect of the extrafocal radiation appears minimized.

A final application that deserves brief mention in this section is the simulation of x-ray units for kV cone beam CT imaging, which is a popular technique used in image-guided radiotherapy (Ay and Zaidi, 2005; Jarry et al., 2006; Mainegra-Hing and Kawrakow, 2008; Downes et al., 2009; Spezi et al., 2009). Figure 5.16 gives an overview of a simulated cone beam CT and the distribution of the last interaction sites of the photons in the geometry (Spezi et al., 2009). These simulation studies may serve to remove scatter from the cone beam images, to optimize the imaging chain, or to derive the patient imaging dose.

### 5.2.6 $^{60}$Co Teletherapy Units

As the relative importance of traditional $^{60}$Co teletherapy in comparison with accelerator therapy has dwindled in industrialized countries, the number of Monte Carlo studies on these modalities is also limited. Although $^{60}$Co decay gives rise to only two gamma-ray lines of 1.17 and 1.33 MeV, the spectrum of photons in a teletherapy unit is affected by the source encapsulation as well as by the collimating system. Contaminant electrons are also an issue, especially in calibration setups. Monte Carlo calculations have been used to characterize $^{60}$Co beams in clinical as well as standard dosimetry laboratory setups.

Han et al. (1987) modeled the Theratron 780 teletherapy machine with a simplified source capsule. Only about 72% of the photons were found to be either 1.17 or 1.33 MeV; the other 28% was single or multiple scattered photons mixed with a small number of unscattered bremsstrahlung and positron annihilation photons. They studied photon fluence spectra, the effect of field size on output, and dose calculations. It was found that an increase in output of the $^{60}$Co unit with increasing field size is caused by scattered photons from the primary tungsten collimator. The electron contamination of an AECL $^{60}$Co teletherapy unit was studied (Rogers et al., 1988) and it was found that electrons originate from the source capsule (most effective at close SSD), the collimators, and the air (most effective at large SSD). At 80 cm SSD, for a field size equivalent to $35 \times 35$ cm$^2$, about 45% of the dose at $d_{max}$ is due to contaminant electrons.

![Figure 5.15](image-url) **FIGURE 5.15** First-generation electron backscatter coefficient $\eta_1$ defined as the ratio of the number of electrons backscattering of an x-ray target, over the number of total primary electrons impinging on it (a). Energy spectra from electrons backscattered at all angles from a tungsten x-ray target (b). (From Ali, E. S., and D. W. Rogers. 2008. Med Phys 35:4149–4160. With permission.)
of which 22% come from the source capsule and air-generated electrons; 10% and 13% are from electrons emerging from the outer and inner collimators, respectively. The major source of scattered photons was the source capsule itself and scattered photons contribute 18% of the maximum dose. The inner and outer collimators were found to increase the effect of photon scatter by a few percent. More recent studies (Mora et al., 1999; Carlsson Tedgren et al., 2010) modeled 60Co units from calibration laboratories in great detail and focused on small but important spectral effects.

Besides its use in traditional teletherapy, 60Co is implicated in special treatment techniques such as the GammaKnife. Of special interest for this treatment modality is the study of dose distributions and output factors for the very small field sizes used for which detailed source modeling is important. Detailed Monte Carlo studies can be found in the literature (Moskvin et al., 2002; Cheung and Yu, 2006).

5.3 Summary

Monte Carlo simulations are an extremely powerful tool in modern radiotherapy. That the Monte Carlo community is alive and well can be seen from successful workshops that were held in recent years in, for example, in Montreal, Ghent, Cardiff, and Sevilla.

Most of the studies cited in this chapter used EGS/EGSnrc-based codes (including the user interface BEAM) but other codes such as GEANT4 and PENELOPE are also gaining popularity. We highlighted the capabilities of these codes for modeling and analyzing external radiotherapy photon beams, but they also have proven to be of great value for kV x-ray systems, brachytherapy sources, and patient dose calculations. In addition, Monte Carlo simulations play an increasingly important role in imaging for diagnostics and patient treatment verification. In this and other chapters, the 4D modeling of linac beam delivery in volumetric arc therapy was mentioned. Elsewhere in this book, the capabilities of Monte Carlo codes to incorporate 4D motion, such as in patient/organ movement and deformation, were discussed.

Simulation time and storage of large amounts of data, once prohibitive for complex Monte Carlo modeling, are now less of an issue. Technology is aiding here, for example, with the advent of ultrafast GPU processors.

In the near future, Monte Carlo-based treatment planning will become the workhorse of patient planning. Speed and accuracy will be of the essence, and clear guidelines on commissioning and model generation will be needed.

It is clear that Monte Carlo methods will remain very important in the study of new technology in radiotherapy. To name just one example, Monte Carlo methods are facing new challenges in, for example, the design and dosimetry in hybrid MRI-linac systems (Kirkby et al., 2008; Lagendijk et al., 2008) where electrons are perturbed by the strong magnetic fields employed in MRI imaging.

References

Ali, E. S., and D. W. Rogers. 2008. Quantifying the effect of off-


6.1 Introduction

The introduction given in Chapter 5 also serves as an introduction to this chapter. In this chapter, we will give an overview of the work performed in the field of Monte Carlo modeling of clinical electron beams over the last 30 years or so. It is widely believed that to calculate a patient dose correctly in electron beams, Monte Carlo simulations are needed, more so than for radiotherapy photon beams where there exist alternative dose calculation methods that rival Monte Carlo simulations in accuracy and performance. For electron beams, the situation is different, with many papers exposing flaws in analytical dose calculation methods, even in modern treatment planning systems. It is, therefore, not surprising that Monte Carlo simulations are seeing a widespread acceptation in the radiotherapy community.

6.2 Electron Beams from Clinical Linacs

Electron beam radiotherapy is administered much less frequently than photon beam treatment, most commonly as a boost to a small superficial target. The energies used nowadays mostly range from 4 to 20 MeV. The schematics of an electron beam linac for Monte Carlo modeling are represented in Figure 6.1. The most important differences with a photon beam linac are the absence of the photon target, the presence of a scattering foil to broaden the beam, and a multistage collimator to shape the beam close to the irradiated volume. The latter is needed because electrons scatter much more in air than photons and they therefore require collimation close to the surface of the patient. Monte Carlo modeling of electron beams will be treated more concisely in this chapter than photon beam modeling because of its less frequent clinical use. The BEAM Monte Carlo code is again the most heavily used tool in many studies.

6.2.1 Early Work in Electron Beam Modeling

Berger and Seltzer (1978) were among the first to model the interaction of electrons with lead scattering foils, which is the linac component that influences the electron beam most significantly and where contaminant bremsstrahlung photon production takes place. They found that the intervening air causes a significant energy degradation of the electron beam, while the effect of the air can be ignored in high-energy photon beams. It needs to be pointed out that the same group that pioneered Monte Carlo photon beam modeling was also one of
the first to present a Monte Carlo model for a clinical electron beam. Borrell-Carbonell et al. (1980) published in 1980 simplified models of several linacs; they modeled beam collimators as apertures with no interactions in their walls; therefore, no secondary particles were produced. This approximation does not result in realistic particle fluence for electron beams. Rogers and Bielajew (1986) compared calculated and measured depth dose curves in water for monoenergetic electrons. They noted that the simulations predicted a less steep dose gradient near the surface and a too steep dose fall-off beyond the depth of maximum dose. When the electrons were made to pass through simulated exit window, scattering foils, and air, the differences were reduced. They pointed out that electron energy straggling has an important influence on the depth dose. Andreo and Fransson (1989) studied stopping-power ratios in water derived from a variety of simple and more complex electron beam models. They showed that stopping powers are relatively insensitive to the details of the electron spectra. They also indicated that it is important to preserve the correlation between energy and angle, which has implications for virtual source models. In another paper (Andreo et al., 1989), they studied the influence of the electron energy and angular distribution on the depth dependence of stopping-power ratios. Figure 6.2 shows that for a variety of simulation geometries, from simple to more complex (depicted in the insert in (a)), the electron energy spectra are more or less broadened. This causes small differences in water depth doses and, in turn, in water to air stopping-power ratios. Ebert and coworkers (Ebert and Hoban, 1995a,b, 1996) studied simple models of applicators and cerrobend cutouts. They identified two main processes by which the applicator/cutout may influence the dose distributions: by electron scatter of the inner edges of the aperture and by bremsstrahlung production. Studies like these may lead to optimized collimator design. Other early studies of simplified Monte Carlo models have been published (Manfredotti et al., 1987; Keall and Hoban, 1994).

6.2.2 Monte Carlo Modeling of Complete Electron Linac Beams

6.2.2.1 First Full Linac Models

In general, it is accepted that modeling electron beams is more difficult than photon beams because of the greater sensitivity of particle fluences and absorbed dose distributions to the details of the primary electron beam (energy, spatial, and angular distribution) and the geometry of the linac, in particular, the scattering foils that are present in most linacs, and the applicators/collimators. The pioneering efforts to model complete electron beam geometries with the EGS4 code were made by Udale/ Udale-Smith (1988, 1992). Figure 6.3 shows her models for Philips linacs, which included the exit window, the primary collimator, the scattering foils, the monitor chamber, the mirror, the movable photon jaws, the accessory ring, and the applicator.

The modular approach of building linacs paved the way for the later BEAM code with its "component module" approach. To quantify the effects of various parts of the treatment head, she simulated five cases, from a monoenergetic pencil beam in vacuum to the full linac geometry. She used measured depth dose distributions to tune the monoenergetic primary electron source energy by attempting to match the depth of 50% of the maximum dose, $R_{50}$, and the practical electron range, $R_p$. Electron range rejection was employed as variance reduction technique to avoid transporting electrons that could not reach the linac exit at great computational cost. Computers have since then become many orders of magnitude faster but it is still good advice to invest time in selecting variance reduction techniques wisely. Udale scored phase-space files at the bottom of the linac, and in a second step these were used for phantom dose calculations, still a common approach today. She also used a second approach, namely, by extracting energy and angular distributions from the phase-space files and using them in a virtual linac model. Correlations between particle position, energy, and angle are ignored this way. She demonstrated that some degree of correlation must be maintained to avoid loss of
simulation accuracy. Another approach to reduce calculation time was to calculate dose in water separately for primary electrons, secondary electrons, and photons and to scale their contributions. Udale-Smith (1992) compared models of several linacs and established that some had superior designs, that is, lead to improved dose distributions, in that they produced less contaminant photons, less energetic contaminant photons, less scattered electrons, and narrower electron angular distributions. Monte Carlo simulation is the ideal tool for these kinds of studies.

Other early work was done by Kassaee et al. (1994) on applicator modeling and by Burns et al. (1995) on electron stopping powers and practical ranges in clinical electron beams.

6.2.2.2 Applications of the BEAM Code

As stated before in the sections on photon beam modeling, the advent of the BEAM code in 1995 (Rogers et al., 1995) was a major step forward for Monte Carlo modeling of linacs, including electron beams. In fact, most early results reported with the BEAM code were on electron beams. Also, for electron beams, a wide variety of geometry modules (“component modules”), source geometries, variance reduction techniques, scoring techniques, and tagging methods are available. It remains a very useful and popular tool nowadays.

Figure 6.4, taken from the original BEAM paper (Rogers et al., 1995), demonstrates the excellent agreement that can be obtained for dose distributions with BEAM simulations, provided all the necessary details of the linac are known. This is often a problem such that the user has to resort to “tuning” the model, but in this case it concerned a research linac with very well-known characteristics. A more complete report including medical linacs was given by Ding and Rogers (1995). An example is given in Figure 6.5, including again a representation of particle tracks interacting with the linac.
In a series of reports, Ma et al. (1997) and Ma and Rogers (1995) investigated the concept of multiple source models that made use of virtual point source positions due to the diffusivity of electron scatter. This makes deriving source models for electron beams more complicated than for photon beams, which exhibit less scatter. Using realistic linac models, they calculated mean energy in water and stopping-power ratios in water (Ding and Rogers, 1996). With a similar approach, electron fluence correction factors used in the conversion of dose measured in plastic to dose in water were also reported (Ding et al., 1997).

Various authors focused on studying components of electron linacs. Verhaegen et al. (2000) studied backscatter to the monitor chamber, an effect potentially present in both electron and photon beams in some linacs. The relative increase in backscatter (increased signal of monitor chamber) was found to be 2% when the photon jaws decrease from a square field size of 40 to 0 cm in a 6 MeV beam. For higher energies, the effect was smaller. The magnitude of the effect is comparable to photon beams. The contribution to the backscattered fluence from downstream linac components other than the jaws (applicator scrapers, field-defining cut-out) was found to be negligible. Figure 6.6 shows the spectra of forward- and backward-moving particles at the monitor chamber. In contrast to photon beams, in electron beams, we obtain a significant difference in spectral shape between the forward and backscattered electrons. As in photon beams, the backscattered photons were found to contribute insignificantly to the energy deposited in the chamber.

**FIGURE 6.5** (See color insert.) Monte Carlo model for Varian electron beam (top). The bottom shows particle energy and angular spectra for a 15 MeV electron beam, differentiated according to their origin or interaction site. (Reproduced from Ding, G., and D. Rogers. 1995. Energy spectra, angular spread, and dose distributions of electron beams from various accelerators used in radiotherapy. Report PIRS-0439 (Ottawa: NRCC). With permission.)
Dose calculations in electron beams are sensitive to the characteristics of the electron spot size hitting the scattering foils. Information on the photon focal spot size in photon beam mode is not helpful for the electron mode, so it must be determined separately. An experimental technique based on a slit camera has been proposed (Huang et al., 2005) to derive the size of the source of bremsstrahlung photons emerging from the scattering foils, which is equivalent to the size of the electron beam impinging on them. They derived a full width at half maximum (FWHM) size of the elliptically shaped electron primary beams as 1.7–2.2 mm for 6–16 MeV electron beams. They also noted a shift of the primary beam up to 8 mm away from the center of the linac. These authors also studied the influence of primary electron beam divergence on dosimetry in large fields. Introducing a 1.6° primary electron beam divergence changed in air measured beam profiles drastically. Primary electron beam information is among the hardest to estimate in clinical electron beams.

It is known that electron beam models are sensitive to the details of the scattering foils, especially for large fields of high-energy electrons. A study (Bieda et al., 2001) considered four parameters of the scattering foils and found that the distance between the foils is critical. This study emphasizes the importance of communication with the manufacturer or verification of the geometry. Large electron fields were also studied in great detail by Faddegon and coworkers (Faddegon et al., 2005, 2008, 2009; Schreiber and Faddegon, 2005; Shea et al., 2010).

### 6.2.2.3 Studies for Electron Treatment Planning

It is known that conventional treatment planning systems for electron beams have errors in irregular fields and heterogeneous targets (Ding et al., 2005). It is generally accepted that Monte Carlo algorithms have unparalleled accuracy for electron dose calculations. Recently, several Monte Carlo-based treatment planning systems became available. These have been evaluated extensively (Cygler et al., 2004, 2005; Ding et al., 2006; Popple et al., 2006), which is covered in another chapter in this book. Much of the present-day work is based on the pioneering efforts of a few groups. Fast Monte Carlo codes with electron treatment planning as their main application were introduced by Neuenschwander et al. (1995) and Kawrakow et al. (1996) in the mid-1990s.

Ma and coworkers considered various aspects of Monte Carlo electron treatment planning. They simulated clinical electron beams (Kapur et al., 1998; Ma et al., 1999) and studied beam characterization and modeling for dose calculations (Ma, 1998; Jiang et al., 2000), air gap factors for treatment at extended distances, the commissioning procedure for electron beam treatment planning, stopping-power ratios for dose conversions (Kapur and Ma, 1999), and output factors (Kapur et al., 1998). The latter were also investigated by others (Zhang et al., 1999; Verhaegen et al., 2001). In most of these studies, the BEAM code was used. Virtual source models for electron beams remain an active field of research (Wieslander and Knoos, 2006, 2007).

### 6.2.2.4 Other Studies with Electron Beams

Several workers developed intensity-modulated electron therapy, in analogy with photon beams. This involves designing or modifying an electron collimator close to the patient and developing optimization methods for the beam delivery. Al-Yahya et al. (2005a,b, 2007) designed a few-leaf electron colimator, consisting of four motor-driven trimmer bars. Monte Carlo modeling was heavily relied upon during the design stage. The full range of rectangular fields that can be delivered with the device, in combination with the available electron energies, were used as input for an inverse planning algorithm based on simulated annealing (Al-Yahya et al., 2005a). This system optimizes the beam delivery based on precalculated patient-specific dose kernels. They demonstrated that highly conformal treatments can be planned this way.

Modulated electron radiation therapy (MERT) whereby both intensity and energy of the beams are modulated was studied intensively by Ma and coworkers (Ma et al., 2000; Lee et al., 2001, 2000; Deng et al., 2002; Jin et al., 2008; Klein et al., 2008). They first did this for a dedicated electron multileaf collimator (Ma et al., 2000; Lee et al., 2001, 2000; Deng et al., 2002), and later by employing the standard photon multileaf collimator to shape the electron beams at short treatment distances (Jin et al., 2008; Klein et al., 2008). They also used inverse Monte Carlo treatment planning.

A final application of electron Monte Carlo modeling to be mentioned in this chapter is the use of portal imaging in electron beams. The bremsstrahlung photon contamination may be used to obtain an image of the patient as was demonstrated by relying on Monte Carlo techniques (Jarry and Verhaegen, 2005) to study the response of the portal imager to the bremsstrahlung photon fluence. Figure 6.7 compares calculated with recorded portal images in an electron beam. It is clear that also here Monte Carlo techniques play an important role in the optimization of the imaging chain.
Monte Carlo Techniques in Radiation Therapy

6.3 Summary

As in the Chapter 5 on Monte Carlo simulations for photon beams, these techniques are a very useful powerful tool in modern radiotherapy with electron beams. This is even more the case here due to known deficiencies of analytical electron dose calculation techniques. In electron beams, Monte Carlo simulations also play a decisive role in designing complex beam delivery techniques such as in modulated electron therapy. Also in this field, studies with the EGS/EGSnrc-based codes (including the user interface BEAM) dominate, but other codes such as GEANT4 are experiencing an increasing use. Since Monte Carlo codes usually differ more in their cross sections and transport methods for electrons than for photons, care has to be taken to properly benchmark simulations for electron beams. Variance reduction techniques for electrons are another point of interest. Many clinical treatment planning systems are offering Monte Carlo modules for electrons, which will be covered elsewhere in this book.

References


7.1 Introduction

Dynamic delivery techniques such as dynamic wedges and intensity-modulated radiation therapy (IMRT) employing dynamic multileaf collimators (MLCs), tomotherapy, and intensity-modulated arc therapy (IMAT) are being used on a routine basis in many radiotherapy clinics. In a dynamic beam delivery, the particle fluence is modulated by a beam modifier whose position within the beam is varied as a function of time. In some techniques, not only the fluence is modified but also the incident direction and energy of the beam may be changed, for example, in an arc therapy delivery or a scanned proton beam. In the absence of Monte Carlo (MC) simulations, the influence of the beam modifiers on the spatial and energy distribution of incident particles is often approximated by the treatment planning system with poorly defined dosimetric consequences, especially when multiple field segments or beam angles are delivered. Accurate dose calculation methods are required to characterize the dynamic beam modifiers, to test the accuracy of the treatment planning system dose distribution, and to perform independent monitor unit calculations as well as to reconstruct the patient dose delivery. If properly validated, a dynamic MC model of the beam can serve as a commissioning tool to replace extensive complicated measurements, especially if measurement resolution or accuracy is questionable. This topic comprises the first part of this chapter. The methods for simulation of dynamic beam delivery devices will be discussed followed by examples of specific applications to dynamic radiotherapy techniques.

The simulation of dynamic patient geometries has only been investigated recently as the research interest in four-dimensional (4D) radiotherapy techniques for management of respiratory motion has increased. The use of MC-based dose calculation algorithms is a natural choice for such applications due to their excellent dosimetric accuracy in low-density lung tissue compared to the more standard analytical dose calculation algorithms. Different approaches to dynamic patient dose accumulation using MC methods will be discussed in the second part of this chapter as well as issues that may limit their practical application.

Finally, the interaction of dynamic beam delivery methods with tumor motion, referred to as the “interplay effect” is a concern in intensity-modulated photon and proton therapy, which employ moving MLCs or scanned beams. Numerical methods are best suited for such studies, including approaches that use MC algorithms. This topic comprises the final section of this chapter.

7.2 Simulations of Dynamic Beam Delivery

7.2.1 Strategies for Simulating Time-Dependent Beam Geometries

In developing MC algorithms to model the IMRT delivery, a very accurate “blueprint” (geometry, atomic composition, etc.) is required of the beam-modifying devices used to generate the intensity profile. The approaches for simulating dynamic geometries in MC simulations range, in the order of increasing complexity, from modification of particle weights to account for attenuation, performing multiple static simulations of ”discrete” geometrical states, to treating the beam motion as a probabilistic problem where the beam geometry is sampled on a particle-by-particle basis from probability distributions describing the
fraction of the total delivery time for which each geometrical configuration exists.

The simplest approach, particle weighting, arose from techniques used in analytical dose calculation algorithms to model the influence of beam modifiers. This approach has been favored for its relative computational efficiency compared to simulating particle transport in the beam modifiers. Weighting factors may be determined from linear attenuation based on a ray tracing through the beam modifier geometry. Temporal variations of the beam modifier positions are accounted by scaling the weighting factors by the fraction of the total delivery time that the modifier blocks the beam path. For example, Ma et al. (2000) converted MLC leaf sequence files to intensity maps by accumulating monitor units (MUs) for unblocked areas and for the blocked areas, where the latter contribution is weighted by the average leaf leakage. This approach ignores the effect of the leaf tongue and groove as well as the scatter contributions from the leaves. To model more accurately the MLC leaf geometry, a ray-tracing algorithm can be used to calculate intensity maps (Deng et al., 2001) to more accurately model the influence of the leaf geometry. An example of the fluence maps calculated with and without consideration of the leaf geometry is shown in Figure 7.1.

Keall et al. (2001) also included the scatter contributions by calculating probability maps for leaves being open, closed, or a leaf tip passing through a grid point. The weighting factors for blocking leaves are calculated from linear attenuation along the path length of the particle through the MLC considering the energy of the incident photon. The scatter contribution is obtained by integration of attenuation factors obtained for successive positions of the leaf tips by subsampling the intensity map grid. These MLC simulations should be distinguished from the other previously mentioned approaches because they also include modeling of first-Compton scattering of each incident photon; thus, a simplified photon transport in the MLC was performed. The weights of these scattered photons are modified by the probability of a leaf blocking the particular grid point. A limitation of the particle weighting approach is that the nonuniformity of particle weights leads to greater statistical variance (Ma et al., 2000). To improve the particle weight uniformity, particle splitting may be applied to particles with large weights, while Russian roulette can be used to reduce the number of low-weight particles.

Simulations that model in detail the different geometrical states of a dynamic beam delivery can be divided into two main approaches. The first uses multiple discrete simulations and will be referred to as the “static component simulation” (SCS) method. This approach is logical when geometry changes occur in discrete steps (e.g., the step-and-shoot IMRT); however, for a continuously variable geometry, a reasonable limit must be imposed on the number of geometry samples and thus the temporal resolution. It should be pointed out that simulating more geometries does not necessarily lengthen the calculation time.

FIGURE 7.1  (See color insert.) (a) and (c) Intensity maps used for particle weighting calculated for one field of a step-and-shoot IMRT plan. On the right, (b and d) are shown as the corresponding Monte Carlo dose distributions. The intensity map in (c) is calculated using a ray tracking algorithm, which includes the tongue-and-groove geometry of the MLC leaves. (From Deng J. et al. The MLC tongue-and-groove effect on IMRT dose distributions. Phys. Med. Biol. 46, 1039–1060, 2001. With permission.)
as the total number of histories to be simulated can be distributed between the individual geometries. Rather, the overhead comes from the input file preparation, initialization, and post-processing steps. These SCS simulations can be run separately and the results can be recombined later or by restarting a previous simulation with an updated geometry (Shih et al., 2001). Furthermore, the simulations on different geometries can be performed for equal numbers of particle histories and the results can be weighted by the fractional MUs to determine the dose. Alternatively, the number of histories to be simulated for each geometry can be calculated based on the fractional MUs. With this latter approach, the statistical variance will be higher for segments that deliver fewer MUs.

The time dependence of the geometry can be incorporated into a fully dynamic MC simulation by continuously updating the geometry after a predefined number of initial particles, corresponding to the time resolution. In GEANT4, this is possible through the use of the G4UI messenger class, which uses the method SetNewValue to update the geometry parameters (Paganetti, 2004). This approach updates the geometry in a linear time progression. Another approach is to randomize the sampling of the different geometries. This “position-probability sampling” (PPS) method (Liu et al., 2001) requires cumulative probability functions (cpdf) for each geometrical parameter that varies. The probabilities are calculated from the fraction of the total delivery time that the particular geometrical element (i.e., MLC leaf and jaw) spends at a certain location or configuration. From an operational overhead point of view, the PPS approach may be more efficient than the SCS method. Both approaches should result in the same statistical variance for an equal number of incident particles if the SCS method calculates the number of histories to run for each geometry based on the same cpdf used for the PPS method.

7.2.2 Applications of Monte Carlo to Model Dynamic Radiotherapy Techniques

7.2.2.1 Dynamic Wedge

Dynamic or virtual wedges are an intensity-modulated delivery approach where a wedge-shaped dose distribution is delivered by the dynamic motion of the collimating jaws. The jaw motion is specified as a function of fractional MUs in the so-called segmented treatment table (STT). The spectrum of particles emerging from a virtual wedge may vary significantly from a static solid wedge, which differentially hardens the beam.

The first reported simulation of a dynamic wedge was performed by Verhaegen and Das (1999). They used the EGS4/BEAM code to simulate delivery with a dynamic wedge on a Siemens linac and to compare the energy spectrum with that obtained with a physical wedge. In the first step, the simulation of transport in the upper section of the treatment head is performed and a phase space file is scored before the upper jaws. In the next step, 20 discrete simulations of transport through the treatment jaws are performed between which one of the upper jaws moved in 1 cm steps. The resulting 20 phase space files are combined by taking a precalculated number of particles from each phase space file based on a formula for the ratio of dose, or MUs, that must be delivered at each jaw position to obtain a certain wedge angle. This assumption of a one-to-one relationship between the number of delivered MUs and the number of simulated histories is not entirely accurate as it ignores the effect of particles that are backscattered from the jaws into the monitor unit chamber. A correction for this can be applied if the amount of backscatter as a function of jaw position can be characterized for the linac geometry (Liu et al., 2000). In the study of Verhaegen and Das (1999), the dose in the monitor unit chamber resulting from backscattered radiation was less than 1% for all simulated jaw positions. The authors modeled both virtual and physical wedges from 15° to 60° for energies 6–10 MV and obtained a good agreement with measurements, except for the penumbra region of the toe end of the wedge where the discrepancies were up to 4%. They also compared the virtual wedge with the physical wedge in the heel, center, and toe end of the wedges for beam spectrum variations. In comparing physical wedges with virtual wedges for beam hardening effects, no major differences were observed except that the 60° physical wedges produce significantly harder beams across the whole field due to higher absorption by the tungsten.

In a similar fashion, Shih et al. (2001) reproduced the dynamic wedge delivery of a Varian linac, as specified in an STT file, by calculating the number of particles to be simulated for each jaw setting based on the weights in the STT. After the simulation of the first STT entry, the simulation is restarted using the IRESTART feature in BEAM with the updated jaw position and incident histories. In this way, the phase spaces scored for each jaw setting are appended to the main phase space file and no additional postprocessing is needed.

Verhaegen and Liu (2001) later developed the PPS method to simulate the Varian enhanced dynamic wedge (EDW) delivery. The jaw positions as a function of cumulative MUs in the STT are converted to a cumulative probability distribution function (see Figure 7.2). During initialization of each incident history, the jaw position is randomly sampled from this cpdf. The authors also performed simulations using the SCS method with and without a correction for the reduction in the number of delivered MUs due to backscattering into the monitor chamber. The incorporation of this correction was not found to make any difference in the phase space files, thus again confirming the validity of the assumption relating MUs to particle histories. Comparing the results of the PPS and SCS methods, no differences were found in the resulting phase space files created with the two methods.

The use of the particle weighting method to simulate a Varian EDW delivery has been reported by Ahmad et al. (2009) using the method of Ma et al. (2000). Compared to measurements of wedged profiles and output factors, they obtained agreement within 2% and 1 mm.

7.2.2.2 MLC-Based IMRT

With IMRT, nonuniform fluence patterns or intensity-modulated maps that are incident upon a patient from multiple
directions are used to generate a uniform dose distribution that covers the target volume. These nonuniform fluence patterns are generated by using a continuous or discrete sequence of apertures formed using an MLC. In step-and-shoot delivery, each MLC-generated subfield is set prior to the beam turning on and there is no MLC motion during irradiation. In the case of the sliding window (or dynamic) delivery, the leaves move while the beam is on.

MC simulations of step-and-shoot and sliding window IMRT have been used by numerous authors to compare MC dose distributions with those produced by a commercial planning system. For example, Ma et al. (2000) used the particle weighting method to recalculate IMRT treatment plans from the CORVUS treatment planning system by using the leaf sequence files to calculate a two-dimensional (2D) intensity map.

Leal et al. (2003) and Seco et al. (2005) performed the full linac and patient simulation (i.e., particle transport through jaws, MLC, and patient), respectively, for the Siemens and Elekta linac using BEAMnrc, where step-and-shoot delivery was modeled using the SCS method. Each beam consisted of 5–15 segments, which were simulated individually with the number of histories proportional to the number of MUs delivered by the individual segment. Both authors performed their simulations on a personal computer (PC cluster). Scripts were developed to automate generation and distribution of separate files for the simulation of each segment phase space and the resulting dose distribution.

In Liu et al. (2001), the authors simulated both dynamic MLC and step-and-shoot deliveries of IMRT fields using the PPS method. First, the leaf positions as a function of cumulative MUs are used to create a probability function for each leaf (see Figure 7.3a). An MC simulation of an IMRT delivery is performed by randomly selecting on the Y-axis of Figure 7.3a an MU_{index} value between 0 and 1. The corresponding MLC subfield is then obtained by converting the MU_{index} (Y-axis) into a segment index (X-axis) using the curve of either step-and-shoot or sweeping leaf curve. If the X-axis value falls between two segment index values, then the leaf positions at this fractional index value will be linearly interpolated from the neighboring segments. Note that the “step-like” appearance of the step-and-shoot curve results from the beam being turned off when the segment index and the leaf positions change. A similar approach was used by Heath and Seuntjens (2003) for simulations of IMRT delivery with the Varian Millennium 120-leaf MLC using BEAMnrc. They modified the standard VARMLC component module, which is available for modeling MLC geometry, to model specifically the new geometry of the

![Figure 7.2](image-url)


![Figure 7.3](image-url)

120-leaf MLC and to allow simulation of IMRT with both the SCS and PPS methods.

A fourth approach to simulating MLC-based IMRT is through a combination of the geometry sampling approaches and simplification of the particle transport in the MLC. The complexity of the leaf geometries typically leads to a high number of geometrical regions through which particles must be tracked. When particle transport in these detailed geometries is performed with low cut-off energies, these simulations can become very time consuming. Depending on the quantity being investigated, the contributions of electrons and multiple-scattered photons may not be important. Therefore, a significant improvement in computational efficiency can be realized by simplifying both the particle transport and the simulation geometry. An example of this is the fast MC code VCU-MLC (Siebers et al., 2002), which uses a single-Compton approximation of photon interactions within the MLC. No electron transport in the MLC is performed and the incident electrons have their weight reduced by the probability of striking an MLC leaf. The leaf geometry is approximated by segmenting each leaf into upper and lower halves, with a correct thickness to model the tongue and groove. Scattered photons are assumed to pass through the current leaf until they exit the MLC; thus, the adjacent leaves are ignored. The VCU-MLC model predicts correctly the field size dependence of MLC leakage, inter- and intraleaf leakage, and the tongue-and-groove dose effect. The model is applicable to both step-and-shoot and sweeping window deliveries. For each incident particle, the weight is modified by randomly sampling MLC positions 100 times from the leaf sequence file and averaging the transmission probabilities for each sample. Compared to the SCS and PPS methods that sample an MLC configuration for each incident particle, fewer source particles are required with the transmission probability weighting approach, which results in a gain in calculation efficiency. The VCU-MLC code is thus very useful to patient IMRT dose calculations because of its accuracy and speed of calculation. The MLC model has also been extended to simulate photon transport through both jaws and MLC simultaneously with the addition of multiple Compton interactions (Seco et al., 2008). Tyagi et al. (2007) have also implemented a similar approach in the DPM code to simplify particle transport in a model of the Varian Millennium MLC; however, they adopted a serial approach to simulating IMRT delivery by using the PPS method to determine the leaf positions for each incident particle.

An application of MC simulations of MLC-based IMRT delivery is to reconstruct the delivered dose based on recorded delivery log files (Fan et al., 2006; Tyagi et al., 2006). For example, on Varian linacs, the MLC leaf positions and beam state (on/off) are recorded every 0.050 s and written to a Dynalog file. The log files can be processed in a manner similar to leaf sequence files. This can be used as a quality assurance method to detect dosimetric errors resulting from errors in positioning of the MLC leaves. Owing to the large number of points within these files, calculation with an analytical algorithm would prove impractical.

7.2.2.3 Tomotherapy

The challenge of simulating tomotherapy, and other delivery techniques that involve dynamic gantry rotation, is that the gantry angle is continuously changing during the delivery. The geometry of a helical tomotherapy delivery is illustrated in Figure 7.4. Tomotherapy uses a rotating fan beam source whose intensity is modulated by a binary MLC. The helical beam delivery is specified in a sinogram file in which the fractional opening time for each leaf is specified at discrete gantry positions. Since the delivery is dynamic, leaf motion is occurring while the source rotates. Depending on the amount of leaf motion, sub-sampling of the sinogram may be needed to accurately simulate the dosimetric effects of the combined gantry and leaf motion. The amount by which the couch translates as a function of the gantry rotation is specified by the pitch. Static models of tomotherapy units without MLC modulation have been developed by Jeraj et al. (2004) and Thomas et al. (2005).

Sterpin et al. (2008) developed a PENELOPE user code called TomoPen for simulating helical tomotherapy delivery. First, a phase space file is created for each of the three jaw settings used during delivery. Then, for each entry in the sinogram file, a projection-specific phase space file is created by adjusting the weights of the particles in the phase space file associated with the appropriate jaw setting. To model the continuous delivery, each entry in the sinogram file is further subdivided into 11 sub-projections for which the leaf openings are linearly interpolated. It should be noted that the treatment planning system used the sinogram entries to calculate the dose distribution, therefore
discretizing the delivery. For all the particles in the jaw-specific phase space, weighting factors are calculated, assuming linear attenuation, by a ray tracing through the MLC leaves. Specifically, the subprojection-specific phase space is calculated by dividing the jaw-specific phase space into 64 regions, one for each MLC leaf, and applying the weighting for leaf attenuation if the current leaf is closed or not adjusting the weight if the leaf is open. Only leaves that are open or adjacent to an open leaf are considered for this step, otherwise, the particles were discarded; thus, the leakage radiation in this situation is not modeled. Each of the resulting subprojection-specific phase space files is then rotated and translated according to the projection angle and couch translation before being applied to the patient geometry. TomoPen has been used to simulate the delivery of commissioning and clinical cases in homogeneous phantoms. A comparison with film and ionization chamber measurements (see Figure 7.5) showed a good agreement within 2% and 1 mm.

Zhao et al. (2008a) also simulated a helical tomotherapy delivery with EGSnrc/BEAMnrc by using the SCS method. They also subdivided each sinogram entry into static MLC subfields. An initial phase space is scored above the MLC. Then, the binary collimator is modeled using the VARMLC component module and a separate full simulation of particle transport in the MLC is performed for each subprojection resulting in a second set of phase spaces. The number of histories for each subprojection simulation was proportional to the calculated leaf opening time. The phase spaces are rotated in the XY plane and the isocenter is modified along the Z direction to simulate the helical delivery as shown in Figure 7.4. This model was used to compare with dose distributions calculated by their treatment planning system for clinical treatment plans (Zhao et al., 2008b).

7.2.2.4 Intensity-Modulated Arc Therapy
IMAT is a form of conventional MLC-based IMRT where the gantry rotates around the patient as the beam apertures are dynamically modified by the MLC. There are a number of different IMAT implementations, which are distinguished by the number of arcs that are delivered and whether the dose rate is variable or constant. The delivery is specified by an MLC leaf sequence file, which is indexed to the gantry rotation angle.

The approach for simulating IMAT delivery is similar to that for tomotherapy. An SCS-based approach to modeling IMAT delivery on an Elekta linac has been described by Li et al. (2001) using EGS4/BEAM. Emulating the dose calculation method used by the treatment planning system, the authors discretize each arc into 5–10° steps for which static simulations of the MLC apertures defined in the leaf sequence file are performed. A simplified geometry is used for the MLC, without modeling the tongue

![Comparison of dose distributions calculated with a convolution–superposition algorithm (C/S) and Monte Carlo (TomoPen) and measured with film (MS EDR-2) and ionization chamber (MS IC) for a commissioning plan in a cylindrical phantom. (From Sterpin E. et al. Monte Carlo simulation of helical tomotherapy with PENEL...](image-url)
and groove. For each phase space, a DOSXYZ calculation on the patient geometry is carried out with the appropriate source angle.

A model of a Varian RapidArc delivery has been developed by Bush et al. (2008), which used EGSnrc/BEAMnrc. They also modeled the gantry rotation as a series of discrete static simulations. However, owing to the significant leaf motion that occurs during RapidArc delivery, it is also necessary to model the leaf motion that occurs between the gantry control points. This is addressed by calculating, for each sequential pair of gantry angles specified in the sequence file, a mean gantry angle. The leaf openings of the adjacent gantry angles are then used to define a pair of control points for a sliding window dynamic multi-leaf collimator (DMLC) delivery at this mean angle, as illustrated in Figure 7.6. The variable dose rate is accounted for by weighting the subsequent dose calculation by the fractional MUs to be delivered by each segment. The authors used the VCU-MLC model (Siebers et al., 2002) to improve the computational efficiency of their simulations. The same group has used the recorded Dynalog files to reconstruct the delivered dose from a RapidArc treatment (Teke et al., 2009).

Simulation of tomotherapy and IMAT delivery using the SCS approach requires the calculation of phase space files for a significant number of geometry samples due to the additional dimension of gantry rotation. Even with automation of the input file preparation and patient dose calculation, the generation of hundreds to thousands of phase space files representing each subprojection is not optimal in terms of storage requirements. Another drawback of using the SCS approach is that the dosimetric influence of the number of subsamples for simulating the continuous delivery must be carefully investigated.

Owing to the continuous motion of the gantry and collimators, the position-probability approach is better suited to the simulation of IMAT and tomotorhapy. With the PPS approach, a randomly sampled monitor unit index at the beginning of each history can be used to determine the collimator settings as well as gantry angle and isocenter translation. These delivery settings can be obtained from the sinogram/leaf sequence file from the treatment planning system or from a Dynalog file recorded during a treatment delivery.

Two PPS-based approaches have been developed for simulating tomotherapy and IMAT using the BEAMnrc code. The first, by Lobo and Popescu (2010), makes use of the option in DOSXYZnrc to use a full BEAMnrc simulation as a particle source (ISOURCE 9). This DOSXYZnrc source has been modified so that the randomly sampled monitor unit index that is used to determine the geometrical configuration of any dynamic component module in the accelerator is passed to DOSXYZnrc, where it is used to sample the gantry and collimator angle as well as couch rotation and translation from lookup tables resembling the sequence files used to specify the MLC leaf motion. On the basis of this source configuration, the coordinates and trajectory of the particle exiting the accelerator are transformed to the patient coordinate system before being transported through the patient density matrix. These two new sources (ISOURCE 20 and ISOURCE 21) are now distributed with the BEAMnrc V4 package. Source 20 uses a phase space, scored below the secondary collimators (e.g., jaws) as an input. This is useful for simulations when the jaw configuration does not change during the delivery. Lobo and Popescu integrated the VCU-MLC code (Siebers et al., 2002) into Source 20 by compiling it as a shared library, which can be accessed by DOSXYZnrc to transport particles through the MLC. The randomly sampled monitor unit index is used to sample the MLC configuration for each incident particle. The other source (ISOURCE 21) consists of a full accelerator model simulation and is useful when there is more than one dynamic component. In source 21, the same monitor unit index that is randomly sampled when a particle is incident on the first dynamic component is used when that particle enters the second dynamic component. This ensures that the motion of the two collimators is synchronized.

For the second PPS-based approach that was developed by Belec et al. (2011), a time variable is stored in the phase space file. This time variable replaces the ZLAST variable, which records the z position of the last interaction of the particle. The time variable is actually the same monitor unit index that is used by all PPS approaches and will be randomly sampled when a particle passes through the first dynamic component module. Unlike the method of Lobo and Popescu, with this approach, the output of the accelerator simulation can be stored in an intermediate phase space file without losing the temporal information. When the phase space file is read by DOSXYZnrc, the time variable is used to sample from the lookup tables specifying the source configuration.

![Figure 7.6](image-url) Illustration of the generation of DMLC segments for the simulation of leaf motion in volumetric arc therapy (VMAT) delivery. For each consecutive set of gantry positions (1 and 2), a mean gantry angle A is calculated and the leaf positions from 1 and 2 are used as control points for a dynamic leaf sequence to be simulated at position A. (From Bush K., Townson R., and S. Zavgorodni. Monte Carlo simulation of Rapid Arc radiotherapy delivery. *Phys. Med. Biol.* 53, N359–N370, 2008. With permission.)
7.2.2.5 Protons

Before the introduction of scanning delivery systems for proton beams, temporal modulation in the form of spinning range modulator wheels was used for modulating the range of an incident proton beam to create a spread-out Bragg peak. In what is perhaps the first published report of a dynamic MC simulation, Palmans and Verhaegen (1998) reported implementing the PPS method in the PTRAN MC code to simulate a range modulator wheel. Paganietti (2004) described a detailed model of the Northwest Proton Therapy Center beam line using GEANT4. Dynamic elements such as the range modulator wheel and beam scanning magnets (see Figure 7.7) were modeled exploiting the feature of GEANT4 to update geometry parameter values during the simulation. The geometry updates were performed in a linear time fashion, with each time step assigned a number of histories. The modulator wheel rotation was simulated in 0.5° steps whereas the scanning magnet settings were updated in steps of 0.02 T. The authors noted that the calculation time for their dynamic MC simulations were virtually unaffected by the number of times that the geometry was updated since this involved only updating a pointer in the memory.

7.3 Dynamic Patient Simulations

7.3.1 Patient Motion in Radiotherapy

The need for 4D patient dose calculation methods was initiated by an interest in compensating for the effects of respiratory motion during treatment planning and delivery (Keall et al. 2004). 4D dose calculation methods are required to calculate the cumulative dose distribution, which is received by the varying anatomy. The effects of motion on the delivered dose distribution were discussed by Bortfeld et al. (2004). These include: (1) blurring of the dose distribution along the path of motion, (2) localized spatial deformations of the dose distribution at deforming organ boundaries or regions of density changes, and (3) interplay between tumor motion and dynamic beam delivery. As will be discussed in the following sections, the blurring effect can be simply modeled with analytical dose calculation algorithms by convolution methods; however, to quantify dose deformation effects, the influence of individual geometric states on the deposited dose distribution needs to be modeled. The methods discussed here are focused on respiratory motion; however, they can be extended to other examples of patient anatomical variations, such as interfractional geometry variations and applications in adaptive radiotherapy, with appropriate modifications.

7.3.2 Strategies for 4D Patient Simulations

7.3.2.1 Convolution-Based Methods

Similar to approaches used for simulating dynamic beam delivery, the cumulative dose can be calculated either by adjusting the particle fluence or from multiple calculations on different respiratory phases. If the dose distribution is assumed to be shift invariant, then the effect of motion on the dose delivered for a large number of fractions can be estimated by convolving the dose distribution with a probability distribution describing the positional variations (Lujan et al., 1999). This models only the blurring effect of motion on the dose distribution, dose deformations, and differential motion that cannot be modeled by convolution. Furthermore, the assumption of spatial invariance of the dose distribution is not valid at tissue interfaces, which can lead to an underestimation of the dose at these locations (Craig et al. 2001; Chetty et al. 2004).

In the reference frame of the patient, patient motion is interpreted as motion of the beam; therefore, the approach of fluence convolution has also been proposed (Beckham et al., 2002; Chetty et al., 2004). This is analogous to the particle weighting approach; however, the \((x, y, z)\) positions and direction cosines of the particles in the phase space file are modified by randomly sampling shifts from a probability distribution function (pdf) describing the respiratory motion. Fluence convolution attempts to overcome the limitation of the shift invariance assumption in the dose convolution method. However, it is also limited by the approximation of the patient as undergoing rigid body motion.

7.3.2.2 Dose-Mapping Methods: COM and Dose Interpolation

The limitation of the convolution approach is that it does not consider the differential organ motion, deformation, and density changes that cause the shape of the dose distribution to locally “deform” during the beam delivery. To accurately account for these effects, the dose distributions calculated on the different respiratory phases (e.g., exhale, 50% inhale, and 100% inhale) need to be calculated, similar to the SCS method of simulating an IMRT delivery, and weighted by the fraction of the total breathing cycle for which they occur. The number of samples of respiratory states needed to reconstruct the cumulative dose for a full respiratory cycle has been investigated by Rosu et al. (2007). They demonstrated no significant differences in dose volume histograms between cumulative doses calculated from 10 respiratory states and calculations based on only the inhale and exhale states. This conclusion is dependent on the treatment plan design; if dose perturbations due to motion are large, then nonlinearities of the motion between inhale and exhale can be expected to have a nonnegligible influence on the cumulative dose distribution.

Unlike dose deposition from a dynamic beam delivery where the reference frame of the dose deposition remains constant, for a dynamic patient simulation, the position and volume of tissue elements is changing (possibly, in concert with dynamic beam delivery changes). The dose deposited in each geometry needs to be mapped to a reference geometry to determine the cumulative dose deposition from irradiation of the multiple variable geometries. This requires knowledge of the geometrical transformation between tissue locations in the reference and current geometrical states. Image registration methods are commonly used to estimate these transformations. Briefly, the transformation is calculated by finding a transformation, which maximizes the similarity of the two images. This similarity may be based on image intensity or previously identified landmarks.
such as points or contours. A wide range of image registration algorithms exist; mainly, they are classified as being rigid or nonrigid. For the types of anatomical changes observed in patient geometries, deformation of the tissue is observed and thus nonrigid or deformable image registration is best suited for this application. A review of deformable image registration algorithms used in medical imaging has been given by Hill et al. Holden (2008), and Xing et al. (2007). It should be obvious to the reader that the accuracy of the mapped dose calculation depends on the accuracy of the image registration. Some discussion of this topic will be made in Section 7.3.2.4.

There are a variety of dose-mapping methods that are used to map the dose from the geometry on which it is deposited to a reference geometry. These are applicable to any dose calculation method. Generally, dose mapping is discussed as applying to two single geometry samples, but it could easily be implemented as a “true 4D” simulation if the patient geometry is updated during the simulation and the mapping is applied to accumulate the dose depositions on the current scoring geometry to the reference dose grid.

The simplest approach to mapping the dose delivered on the target geometry to the reference geometry is to use the deformation map to determine the voxel in the target geometry that corresponds with the voxel of interest in the reference geometry. The reference voxel is assigned a dose value equal to the dose computed in the target voxel in which the transformed center of mass of the reference voxel lies (see Figure 7.8a). This “center-of-mass (COM)” remapping method ignores contributions in other voxels, which may be overlapped by the transformed reference voxel. This can easily occur if the transformed reference voxels do not exactly overlap the target voxels. Also, if there are any volume changes, then this exact voxel-to-voxel mapping is incorrect. In an application of the COM method, Paganetti et al. (2004) used the updateValue method in GEANT4 to update the patient geometry to one of the 10 different phases of the patient’s breathing from a 4D computed tomography (CT) data set. To accumulate the dose, pointers in the dose matrix were modified using the COM method to ensure that dose depositions on the different geometries were accumulated in the corresponding voxels of the reference geometry.

To deal with possible deformations and the inexact overlap of transformed and target voxels, a trilinear interpolation of the dose from the local neighboring voxels at the transformed COM point in the target geometry can be used to assign the dose in the reference geometry (Figure 7.8b). Voxel volume changes can also be accounted for by subdividing each reference voxel, estimating the mapped dose for each subvolume, and then interpolating these values to assign the dose in the reference geometry (Figure 7.8c). Rosu et al. (2005) reported the application of these methods to dose distributions calculated using an MC algorithm, though the method is applicable to a dose distribution calculated with any algorithm. Dose interpolation methods are computationally efficient; however, like any interpolation method, they suffer from a reduced accuracy wherever dose gradients exist. Heath and Seuntjens (2007) showed that due to these interpolation errors, the dose interpolation method calculates the dose incorrectly in the regions of large dose and deformation gradients. Siebers and Zhong (2008) also demonstrated that in the situation where voxels of the target geometry merge in the reference geometry (i.e., compression), dose interpolation methods do not conserve the integral dose and thus lead to errors in the regions of dose and density gradients.

7.3.2.3 Voxel Warping Method

Two MC approaches to dose remapping that ensure energy conservation have been developed by Heath et al. (2007, 2009), Heath and Kawrakow (2009), and Siebers and Zhong (2008). The first method, called voxel warping, entails deforming the nodes of the reference dose grid voxels using the deformation vectors obtained from registration of the reference to the target image, and was first implemented in a modified version of the DOSXYZnrc user code, called defDOSXYZnrc (Heath et al., 2007). Particle transport is performed on this deformed voxel geometry. As the voxel indices do not change between the reference and deformed state, the energy mapped between these states is conserved. The densities of these deformed voxels are adjusted according to the volume change to conserve mass. To describe the deformed voxels, the faces of the rectangular voxels are divided into two subplanes, forming dodecahedrons (Figure 7.9). This voxel face subdivision leads to a twofold increase in the number of distance-to-voxel boundary calculations that need to be performed by the boundary checking algorithm. Furthermore, because of ambiguities in the calculation of particle–plane intersection points, which assume that the plane has an infinite extent, these intersections need to be further tested to determine if they lie within the boundaries of each plane. These additional computations result in up to a 10-fold increase in computation times with defDOSXYZnrc compared to DOSXYZnrc.

A recent reimplementation of the voxel warping method in VMC++ (Heath and Kawrakow, 2009) achieved a 130-fold improvement in computational efficiency compared to defDOSXYZnrc by using a fast MC code and optimized geometry definitions based on tetrahedral elements. Each deformed voxel is divided into six tetrahedrons, which are used for particle tracking while dose deposition is scored in the voxels. The main advantage of using tetrahedral volume elements is the elimination of the plane–particle intersection ambiguity; furthermore, there are fewer planes to be checked in each tetrahedron compared to the dodecahedron geometry. Compared to VMC++ calculations in rectilinear geometry, the calculation time with the deformable VMC++ geometry is increased by a factor of 2 for the same level of statistical variance. The deformable geometry was implemented as a new geometry class for VMC++, which was compiled as a dynamic shared library to be loaded at run time. Two alternate definitions of the tetrahedral geometry were implemented, differing by the direction in which the voxel faces fold. Though it could be demonstrated that the geometry definition affects the dose deposition, no differences in patient dose distributions calculated with the two alternate geometries could be found (Heath and Kawrakow, 2009).

A possible limitation of the voxel warping method is that it requires a continuous deformation field, which is not always
Dynamic Beam Delivery and 4D Monte Carlo

At the boundary between the lung and the chest wall tissue, sliding occurs, which, at the resolution of the dose calculation and image registration, appears as a discontinuous motion. Before using the displacement vector maps obtained from a deformable image registration, they must be checked for discontinuities in the region where the dose will be calculated. This can be performed by calculating the determinant of the Jacobian matrix of the transformations in the neighborhood of each node of the reference voxel grid (Heath et al., 2007; Christensen et al., 1996). A negative value of the determinant (see Equation 7.1) at a particular node indicates a discontinuity in the transformation \((u_x, u_y, u_z)\) at this node. These can be removed by smoothing the transformations either globally or in the vicinity of the discontinuous node.

FIGURE 7.8 Illustration of different dose-mapping methods. The left grid represents the reference dose grid. On the right side is the dose grid from the irradiated geometry. The dose deposited on the source irradiated geometry is to be mapped to the reference geometry for accumulation. The shaded polygon represents the boundaries of the reference voxel after the reference-to-target geometry transformation is applied. (Parts of this figure adapted from Rosu M. et al. Med. Phys. 32, 2487–2495, 2005.) (a) COM mapping. (b) Dose mapping with trilinear interpolation. (c) Method (b) with subdivision of the reference voxels.

FIGURE 7.9 Deformed voxel geometry used in defDOSXYZnrc obtained by applying deformation vectors to reference voxels.

\[
\det \left[ f(\bar{\chi}) \right] = \begin{vmatrix}
\frac{\partial u_x}{\partial x} + 1 & \frac{\partial u_y}{\partial y} & \frac{\partial u_z}{\partial z} \\
\frac{\partial u_x}{\partial x} & \frac{\partial u_y}{\partial y} + 1 & \frac{\partial u_z}{\partial z} \\
\frac{\partial u_x}{\partial x} & \frac{\partial u_y}{\partial y} & \frac{\partial u_z}{\partial z} + 1
\end{vmatrix}
\] (7.1)
7.3.2.4 Energy-Mapping Methods

Siebers and Zhong (2008) proposed an alternative energy-mapping-based approach in which particles could be transported in the rectilinear target geometry and the energy deposition points are mapped according to the deformation vectors to the target dose grid. To account for the possibility that particle steps in the target geometry may be split over two or more voxels if mapped to the reference geometry, the energy deposition location is randomly sampled along the step. This implementation, termed etmDOSXYZnrc, results in a 10–50% increase in computation time, depending on how the deformation vectors are interpolated, compared with the standard DOSXYZnrc calculations. This approach also does not require a continuous deformation vector field.

However, if the transformation between the reference and target geometry is not exact, the energy mapped to the voxels of the reference phase will not be consistent with the mass assigned to those reference voxels. This means that in the lung, for example, energy could be deposited in a voxel with a density of 1.0 g/cm³, but due to registration error, that energy deposition is mapped to a reference voxel with a density of 0.3 g/cm³. These inconsistent energy–mass mappings lead to a discontinuous dose distribution on the reference geometry. In deformable registration of patient images, determination of an exact transformation is not possible due to a lack of one-to-one correspondence between images arising from image artifacts, noise, partial volume effects, among other causes (Crum et al. 2003). In a further paper, Zhong and Siebers (2009) proposed methods to map the mass along with the energy to ensure that mass is also conserved; thus, the energy and mass distributions used to calculate the final mapped dose distribution are mapped in a consistent manner. This approach is termed the “energy and mass congruent mapping” (EMCM) method. The target voxels are subdivided into 100 subvoxels whose masses are mapped to the reference geometry using the same deformation vectors used for mapping the energy deposition points. The precision of the mass mapping was determined to be 99.95% on a simple deformable phantom geometry. Alternatively, a mass “pulling” approach based on
the reference–target geometry transformation can be used. This requires a method to invert the deformation field. An energy-mapping method, which performs the energy and mass mapping in the reference–target direction is described below. Dose distributions mapped between the inhale and exhale phases of a lung cancer patient using the EMCM method and dose interpolation are shown in Figure 7.10. The mean difference between the EMCM and trilinear dose interpolation (TDI)-mapped dose distributions is 7% of the maximum dose.

Heath et al. (2009) proposed an alternative energy-mapping approach that can be performed, postcalculation, on a dose distribution. Instead of a point-by-point energy remapping, the energy mapped from the target geometry to each reference voxel is determined by the volume overlap of the deformed reference voxels on the target geometry. Each reference voxel is subdivided into tetrahedrons, which are deformed using the displacement vectors, and the volume intersected by each of these tetrahedral and the target voxels is calculated. The energy mapped to each reference voxel is simply the sum over all its tetrahedra of the energy deposited in each overlapped target voxel multiplied by the fractional volume overlap. This second method can be applied to a dose distribution calculated with any algorithm; however, information about the spatial distribution of energy depositions within a voxel is lost. This leads to a loss of accuracy when the energy deposited in one voxel is mapped to multiple voxels on the reference geometry.

### 7.4 Combining Dynamic Beam and Patient Simulations

Combining the simulation of dynamic beam deliveries and variable patient geometries is of interest to study the interplay effects, which arise from the relative motion between the MLC leaves and the moving target region and anatomy upstream from the target, for the case of dynamic MLC delivery. We designate this by “4D-interplay” effects and they were initially studied by Yu et al. (1998). They demonstrated that when IMRT beams are delivered with dynamic collimation, the problem of intrafraction motion causes large errors in the locally delivered photon dose per fraction due to motion in the penumbra region of the beam. The magnitude of the photon dose variations was shown to be strongly dependent on the speed of the beam aperture relative to the speed of the target motion and on the width of the scanning beam relative to the amplitude of target motion. In the case of proton therapy beams, interplay effects may also occur for proton beam scanning where the dose distribution is affected by (i) patient breathing, (ii) proton energy change times, (iii) motion amplitude, and (iv) proton beam rescanning methodology used (Seco et al., 2009a).

Modeling of interplay effects with MC can only be done if every particle is tagged with information that indicates both the beam delivery (e.g., MLC leaf or beam spot positions) through which it travels and the patient geometry state (e.g., breathing phase) to which it will deliver the dose. Particle tagging is not straightforward to setup within an MC simulation because it requires the additional inputs of the (i) beam delivery sequence and (ii) patient breathing pattern. Interplay may be simulated using a motion convolved dose kernel; however, this does not account properly for interface effects or rapid changes in electron density within a specific tissue such as within the lung.

An MC method to study the dosimetric effects of interplay and delivery errors due to dropped segments, dose over/under shoot, faulty leaf motors, tongue-and-groove effect, rounded leaf ends, and communication delays in phantoms has been developed by Litzenberg et al. (2007). Real-time information about the tracking of the MLC leaf positions is obtained from the Dynalog MLC log. Real-time target volume position information is measured using wireless electromagnetic transponders (Calypso Medical Technologies, Inc), where three transponders were implanted in the target allowing position updates 10 times per second. The authors termed this method the “synchronized dynamic delivery” approach. In the MC simulation, each particle is transported through a selected MLC field segment and into a specific breathing phase of the dose cube. Therefore, all breathing phase dose cubes would have to reside in the central processing unit (CPU) memory to reduce input/output (I/O) exchanges. Although the study was performed in phantoms, it was a proof of principle that MC could be used to study interplay effects. However, in the case of patients, the method proposed would require large amounts of CPU memory to work efficiently because of the significantly larger patient CT data cubes.

MC modeling has been used for patients with lung cancer to study interplay effects (Seco et al., 2009b) with a more CPU-efficient method of tracking photons through MLC leaf segments and for a specific breathing phase. To reduce the memory requirements, each photon can be tagged with both MLC segment and the breathing phase to which it would deliver the dose. Photon tagging involves changing one or more bits of the particle information, which is written to the phase file. This can be done readily in codes such as BEAMnrc by altering the LATCH bits to include MLC and breathing phase information. Then, dose calculations can be performed separately on each CT cube using only the particles that were delivered on that phase. Preliminary MC results using this approach showed variations in the delivered dose of 5–8% per beam angle. However, the final dose plan for the lung patient made up of 30 or more fractions was shown to have small or negligible interplay dose effects due to the averaging that occurs over the multiple deliveries.

### 7.5 Summary and Outlook

Three main approaches exist to simulating dynamic beam and patient geometries: particle weighting/convolution, SCS, and PPS. Dynamic simulations of IMRT deliveries is a well-established technique compared to 4D MC calculations in deforming patient geometries. For the latter, dose interpolation methods have been generally adopted for dose accumulation in radiotherapy...
planning; however, the development of accurate 4D MC methods that conserve mass and energy have highlighted the inaccuracies of the interpolation approach.

While particle weighting approaches may appear to be more computationally efficient, comparable performances can be achieved with full “4D” MC simulations using the SCS or PPS method when fast MC codes, with appropriate variance reduction techniques and simplifications of particle transport, are used. The advantage of MC codes compared to analytical algorithms in this regard is that the computation time does not increase as a function of the number of geometries that are simulated since the number of simulated particles may be split between geometries; therefore, a higher temporal resolution can be achieved with MC simulations, which may be of importance for continuous delivery methods such as tomotherapy and IMAT.

The study of temporal interplay effects, which occur when both treatment beam and tumor motion occur is an ongoing issue particularly in proton radiotherapy. MC methods are well suited to examining these problems though an effort has to be made to reduce the CPU memory requirements. Some initial results have demonstrated that this is possible, which opens up further possibilities of using MC as a method to evaluate the magnitude of these interplay effects for different dynamic beam delivery techniques. Another important application is the ability to track deviations of the beam delivery and patient motion from their planned values through reconstruction of the dynamic dose delivery to the patient based on delivery and patient log files.

In summary, time-dependent treatment planning and delivery are being ever more exploited and scrutinized in radiotherapy. MC codes are well suited to the study of temporal effects and should play an important role in their accurate quantification.

References


8.1 General Introduction

In the most general form, the Monte Carlo (MC) method is a statistical approach to solving mathematical and/or physics problems in the form of integration, derivation, and so on. In the case of radiation therapy of cancer, MC is ideally suited to solve the complex problem of particle transport and energy deposition within a heterogeneous medium, that is, the human body. An extensive review of the MC method in radiotherapy is given by Chetty et al. (2007) and references within. In this chapter, we will provide a review of the most widely used MC codes for patient dose calculation with emphasis on the statistical uncertainties, denoising and smoothing methods, Hounsfield unit (HU) to medium conversion, deformable registration, and inverse planning with MC.

There is a large variety of MC codes available now in radiation therapy of which the most widely used are EGSnrc (Kawrakow 2000), MCNP (Brown, 2003), GEANT4 (Agostinelli, 2003), and PENELOPE (Baro et al., 1995; Salvat et al., 2009). All these algorithms are based on the condensed history techniques first developed by Berger et al. (1963), where a large number of electron interactions are condensed into one because electrons lose very little energy in a single electromagnetic interaction. Condensed history implementation was divided into two main classes. In the class I condensed history approach, all collisions are subject to grouping, with the effect of secondary particle creation above a specified threshold energy taken into account after the fact (i.e., independently of the energy actually lost during the step). This is performed by transporting a number of secondary particles. MCNP is an example of a class I type MC code, where the system adopted the electron transport algorithm from ETRAN (Berger and Seltzer, 1973; Seltzer, 1988). This algorithm was developed at NIST by Berger and Seltzer following the condensed history techniques proposed by Berger (1963). In the class II condensed history approach, interactions are divided into “hard” (also designated as “catastrophic”) and “soft” collisions. Soft collisions are grouped in class I scheme, while for hard collisions, the particles are tracked individually as independent particles. EGSnrc, GEANT4, and PENELOPE are examples of class II MC simulations. An important aspect of condensed history is the simulation of a particle crossing a boundary between different regions. A lot of work has gone into MC research of boundary crossing and dependency of calculated results on step size. Step-size artifacts and boundary crossing issues are now well understood (Larsen, 1992; Kawrakow and Bielajew, 1998; Kawrakow, 2000) and have led to a significant improvement in the accuracy of MC code predictions. For further details of condensed history techniques, consult Chetty et al. (2007) and the literature within.

8.2 Statistical Uncertainties in Patient Dose Calculation

8.2.1 Introduction

The use of MC simulation techniques in radiation therapy planning introduces unavoidable statistical noise in the calculated...
Thus increasing its efficiency. “Tricks” to accelerate the overall MC algorithm calculation time, VRT methods use “clever” physics, mathematics, or numerical value without changing the number of iterations used.

The performance of an MC calculation is characterized by its efficiency, $\varepsilon$, which combines the estimated variance with the number of iterations, $N$, required to achieve that variance, within a given amount of time. In other words, the efficiency of an MC simulation (also called “figure of merit”) is an indicator of how “fast” the MC code is achieving a specific $\sigma^2$ value. It is defined as

$$\varepsilon = \frac{1}{\sigma^2 N T}, \quad (8.1)$$

where $\sigma^2$ is an estimate of the true variance on the quantity of interest and $T$ is the central processing unit (CPU) time required to achieve this variance. The efficiency of an MC algorithm can be improved by either decreasing the $\sigma^2$ value for a given $T$, or by decreasing $T$ for a given $N$, while not changing the variance. Variance reduction techniques are used to reduce the $\sigma^2$ value without changing the number of iterations used, $N$. The VRT methods use “clever” physics, mathematics, or numerical “tricks” to accelerate the overall MC algorithm calculation time, thus increasing its efficiency.

### 8.2.2 Dose-Simulating Geometries and Calculation of Uncertainties

Different types of scoring geometries are used for three-dimensional (3D) dose calculations, one cubic or parallelepiped voxels being the most common. Moreover, the dose is usually scored in the geometrical voxels that are based on the computed tomography (CT) information. One problem that arises is that decreasing the CT resolution enormously increases the memory usage. One alternative that is being currently adopted is to decouple the scoring grid from the geometrical grid. In the following paragraphs, some examples of scoring geometries are presented, followed by some methods of uncertainty calculation.

#### 8.2.2.1 Dosels

In the PEREGRINE MC dose engine, the particle transport is performed in a patient transport mesh, which is a Cartesian map of material composition and density, obtained from the patient’s CT images. The dose is scored in an array of overlapping spheres that are independent of the material transport mesh, termed dosels. During the dose calculation, the standard deviation in the dosel receiving the highest dose is tracked and when it reaches a level specified by the user, the simulation is terminated.

#### 8.2.2.2 Kugels

The electron macro Monte Carlo (MMC) dose engine calculates the dose distribution from predetermined electron histories that are stored in look-up tables, which were obtained in spherical volume elements, termed kugels, of different materials. The concept was initially introduced by Neuenschwander et al. (1995) for simulating electron beam treatment planning. The MMC algorithm uses results derived from conventional MC simulations of electron transport through macroscopic spheres of various radii and consisting of a variety of media. On the basis of kugel spheres, the electrons are transported in macroscopic steps through the absorber. The energy loss is scored in the 3D dose matrix. The transport of secondary electrons and bremsstrahlung is also accounted for in the model.

#### 8.2.2.3 Segmented Organs

MC dose scoring can be significantly improved if many of the voxels are combined into a large volume, that is, an organ. In this case, we lose the spatial resolution of the dose distribution, but we gain in the ability to predict organ dose very rapidly and accurately with MC. However, contouring CT volumes into organs can be very subjective and depends significantly on the clinician performing the contouring. Therefore, segment organ dose scoring can only accurately be applied in controlled simulation environments such as the NCAT phantom (Segars, 2001). The NCAT anthropomorphic computational phantom is based on the Visible Human dataset (Spitzer and Whitlock, 1998). The phantom uses nonuniform rational B-spline surfaces to represent 3D human anatomy and also allows the incorporation of four-dimensional (4D) motion of the cardiac and respiratory motions. The anatomical parameters of the phantom are set using an input file so that the anatomy parameters can be scaled to dimensions and values different to that of the default settings based on the Visible Human. Such details include the level of desired anatomical detail for the blood vessels and lung airway. The dynamic nature of the phantom is invoked by specifying whether respiratory and/or cardiac motion is to be included.

MC simulations using NCAT have been performed allowing both voxel scoring and organ scoring by Riboldi et al. (2008). A 4D MC framework was developed to study dose variations with motion, within a controlled environment of a phantom. Results show that the MC simulation framework can model tumor tracking in deformable anatomy with high accuracy, providing absolute doses for intensity-modulated radiotherapy (IMRT) and conformal radiation therapy. Badal et al. (2009) developed a realistic anatomical phantom based on NCAT on which MC simulation may be performed for imaging applications, from image-guided treatments to portal imaging. In this case, the organs’ scoring volumes were segmented into triangular meshes.

#### 8.2.2.4 Voxel Size Effects

The scoring voxel size has a large impact on the calculation time because of the amount of time spent in geometry boundary checks and scoring. Cygler et al. (2004) showed that the number
of histories to be simulated, per unit area, is linearly proportional to the mass of the scoring voxels.

### 8.2.2.5 Concept of Latent Variance

There are generally two sources of statistical uncertainty in the 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. (2001) coined the term “latent variance” to describe the uncertainty due to the statistical fluctuation in the phase space generated from the simulation of the treatment head as opposed to the uncertainty due to the random nature of the dose calculation by MC techniques.

The statistical uncertainty in the dose calculated in a phantom by reusing the particles from the phase space file, and assuming that the particles are independent and ignoring the 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 used.

More importantly, 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 for the random uncertainty from the patient calculation. Should latent variance be a significant factor in the total uncertainty, more independent phase space particles need to be used in the patient simulations.

### 8.2.2.6 Batch Method

In the batch method, the estimate of the uncertainty, $s_T$, of a scored quantity, $X$, is given by

$$s_T = \sqrt{\frac{1}{N(N-1)} \sum_{i=1}^{N} (X_i - \bar{X})^2},$$

where $N$ is the number of batches, usually 10, $X_i$ is the value of $X$ in batch $i$, and $\bar{X}$ is the mean value of $X$ evaluated over all batches. The sample size, $N$, is thus given by the number of batches.

As pointed out by Walters et al. (2002), three main problems can be identified with this approach for the uncertainty calculation:

1. A large number of batches must be used or else there will be significant fluctuations in the uncertainty itself because the sample size, $N$, is quite small.
2. Arbitrary grouping of histories into batches ignores any correlations between incident particles.
3. The batch approach adds an extra dimension to the scoring quantities of interest.

### 8.2.2.7 History-by-History Method

The history-by-history method, implemented accordingly with Salvat’s approach (Salvat et al., 2009), allows the elimination of the above-mentioned problems when using the batch method to estimate the uncertainty in the MC dose calculation. The scored dose data are stored in a very efficient way, while the simulation is running. In this method, $X_i$ now represents the scored quantity in history $i$ rather than in batch $i$, and $N$ is the number of independent events. Equation 8.2 can then be rewritten as follows:

$$s_T = \frac{1}{\sqrt{N(N-1)}} \left( \sum_{i=1}^{N} X_i^2 - \frac{\sum_{i=1}^{N} X_i}{N} \right),$$

(8.3)

When using phase space sources to calculate the dose distribution, one history is defined to be all particle tracks associated with one initial particle. It should also be emphasized that the quantity $X_i$ may be weighted quantities, if variance reduction techniques are used.

During the MC calculation, that is, on the fly, a record of the quantities $\sum_{i=1}^{N} X_i^2$ and $\sum_{i=1}^{N} X_i$ is kept. At the end of the calculation, the uncertainty of the calculation is determined according to Equation 8.3, without the need to store the scored quantity in batches. More details on the algorithm implementation can be found in Sempau et al. (2000) and Salvat et al. (2009).

### 8.3 Denoising and Smoothing Methods

#### 8.3.1 Introduction

As previously mentioned, one possible way to reduce the statistical fluctuations is by performing the MC simulation for a very large number of histories, which is generally not practical without the application of powerful variance reduction techniques. Another possibility is the process of removing or reducing the statistical fluctuations from a noisy calculated data with smoothing or denoising algorithms that are routinely applied in a variety of disciplines that deal with noisy signals (e.g., imaging), thus speeding up the MC calculations.

Analogous to image restoration problems in the field of image processing, an improved estimate of the true image can be produced by smoothing or denoising the raw MC result, produced with fewer source particles than needed with the denoising process resulting in an accelerated image of the true dose distribution. Obviously, the goal of MC denoising algorithms is to be as aggressive as possible in locally smoothing the raw MC result while attempting to avoid the introduction of systematic errors—bias—especially near sharp features such as beam edges.

The work of Keall et al. (Keall, 1999, 2000), Jeraj and Keall (Jeraj et al., 1999), and Buffa et al. (Buffa, 1999) have been seminal in the study of the effect of statistical uncertainty on the evaluation of MC plans and the possibility of removing these effects on MC calculated DVHs.

Various methods related to image digital filtering, wavelet thresholding, adaptive anisotropic diffusion, and denoising based on a minimization problem have been proposed to solve
this problem. These algorithms are an approximate efficiency-enhancing method since they can introduce a systematic bias into the calculation. Nonetheless, denoising techniques are useful as they can reduce the overall (systematic and random) uncertainty when the random component decreases more than the systematic component increases. However, it must be emphasized that the denoising techniques require proper validation under the full range of clinical circumstances before they can be used with MC dose algorithms.

Kawrakow (2002) introduced five accuracy criteria, which have become a standard, to evaluate the performance of smoothing algorithms, where smoothed and unsmoothed MC calculations must be compared to a “benchmark result” that has been obtained through the simulation of a large number of histories, namely

1. Visual inspection of the isodose lines, where the differences between the benchmark and the smoothed results should be minimal.
2. DVHs. The difference area between two DVHs, one corresponding to the smoothed data and another to the benchmark data, should be quantified. This difference should be small.
3. Maximum dose difference. The maximum dose difference to the benchmark for smoothed and not smoothed MC simulations should be obtained.
4. Root-mean-square difference (RMSD). This is a standard quantitative measure for the degree of agreement between two distributed quantities. The RMSD in an MC calculation without smoothing is solely due to the statistical uncertainties and, after smoothing, it will contain some systematic bias introduced by the denoising process. RMSD should be obtained between the benchmark and the smoothed and unsmoothed dose distribution.
5. $x\%/y\text{ mm}$ test. The fraction of voxels with a smoothed dose value that differs more than $x\%$ from a benchmark dose value at the same point and there is no point in the benchmark dose distribution that is closer than $y\text{ mm}$ to the point that has the same dose value (Van Dyke, 1993). The current accuracy recommendation is $2–3\%/2\text{ mm}$ (Fraass et al., 1998).

### 8.3.2 Denoising Integrated Dose Tallies

#### 8.3.2.1 DVH Denoising Methods

Two groups have introduced techniques for denoising the DVH: Sempau and Bielajew (2000) and Jiang et al. (Jiang, 2000). Sempau and Bielajew (2000) used a deconvolution method, where the calculated DVH was considered as the “true” DVH, which would be obtained after an infinite number of histories, thus yielding zero variance, convolved by noise. The implications of this approach would be that decisions based upon DVHs could be made in a faster time and, more importantly, inverse treatment planning or even optimization methods could be employed using MC dose calculations at all stages of the iterative process, since the long calculation times at the intermediate calculation steps could be eliminated.

In the work presented by Jiang et al. (Jiang, 2000), the MC calculated DVH is treated as blurred from the noiseless and “true” DVH. The technique described is similar to image restoration, where a deblurring function is employed to obtain the noiseless DVH. An estimate of the smoothed DVH is blurred and the difference between this image and the blurred and original image is minimized with a least-square minimization method iteratively.

#### 8.3.3 Dose Distributions Denoising Methods

Denoising of DVHs cannot completely address the problem of statistical fluctuations as the dose distributions are frequently assessed and represented by other means (e.g., isodose lines, calculation of the tumor control, and/or normal tissue complication probabilities). This makes the removal of statistical uncertainties from the dose distribution itself more desirable and relevant.

#### 8.3.3.1 Deasy Approach

The first to propose denoising of the 3D dose distribution was Deasy (2000), who suggested that MC results can be obtained from an infinite number of source particles and a noise source due to the statistics of particle counting. He proposed several digital filtering techniques to denoise electron beam MC dose distributions and demonstrated that these techniques improve the visual usability and the clinical reliability of MC dose distributions.

MC calculated dose distributions with high statistical uncertainty were subjected to various digital filters and a comparison was then performed between the resulting denoised distributions to calculations with much lower uncertainty using isodose representations. The concluding remarks were that smoothing with digital filters was a promising technique for dealing with the statistical noise of MC simulations. The main objection against the use of denoising for MC calculated dose distributions results from the concern that smoothing may systematically alter its distribution.

#### 8.3.3.2 Wavelet Approach

Deasy et al. (2002) presented another approach, where wavelet threshold denoising was used to accelerate the radiation therapy MC simulations by factors of 2 or more. The dose distribution is separated into a smooth function and noise, represented by an array of dose values that are, subsequently, transformed into discrete wavelet coefficients. Below a positive threshold of the wavelet coefficient values, they were set equal to zero. They showed that with a suitable value of this threshold parameter, the statistical noise was suppressed with little introduction of bias.

#### 8.3.3.3 Savitzky-Golay Method

Kawrakow (2002) presented a 3D generalization of a Savitzky–Golay filter with adaptive smoothing window size, that is, the number of surrounding voxels that is used, to reduce the
probability for systematic bias. The size of the smoothing window is based on the statistical uncertainty in the voxel that is being smoothed. According to the author, this filter decreases the number of particle histories by factors of 2–20, concluding that smoothing is extremely valuable for the initial trial-and-error phase of the treatment planning process.

8.3.3.4 Diffusion Equation Method

Miao et al. (2003) investigated the denoising of MC dose distributions with a 3D adaptive anisotropic diffusion method. The conventional anisotropic diffusion method was extended by, adaptively, modifying the filtering parameters according to the local statistical noise. The work presented showed that this method can reduce statistical noise significantly, that is, 2–5 times, corresponding to a reduction in the simulation time by a factor of up to 20, while maintaining the characteristics of the gradient dose areas.

8.3.3.5 IRON Method

Fippel and Nusslin (2003) introduced the so-called iterative reduction of noise (IRON) algorithm, which iteratively reduces the statistical noise for MC dose calculations. By varying the dose in each voxel, this algorithm minimizes the second partial derivatives of the dose with respect to the three spatial coordinates. This algorithm requires as input MC dose distributions with or without known statistical uncertainties.

Smoothing of the MC dose distributions using IRON was proven to lead to an additional reduction of the MC calculation time by factors between 2 and 10. As with the application of the other smoothing techniques, this reduction is particularly useful if the MC dose calculation is part of an inverse treatment planning calculation.

8.3.3.6 Content Adaptive Median Hybrid Filter Method

In this method, linear filters were combined through a weighted sum, with the median operation to produce hybrid median filters (El Naqa et al., 2005). In regions with strong second derivative features, the median operator is used and in smoother regions, the mean operator is preferred. In general, median filters outperform the moving average and other linear filters in the process of removing impulsive noise (outliers), and in the preservation of edges, however, they fail to provide the same degree of smoothness in homogenous regions. An adaptive combination with mean value linear filters was then chosen in this study.

8.4 CT to Medium Conversion Methods

8.4.1 CT Stoichiometric Conversion Methods

The conversion of CT HUs into material composition and mass density may strongly influence the accuracy of patient dose calculations in MC treatment planning or in any other dose calculation algorithm. To establish an accurate relationship between CT HU and electron density of a tissue, the CT scale is usually divided into a number of subsets. The initial MC dose algorithms used six or fewer materials to define the conversion, for example, air, lung, fat, water, muscle, and bone (DeMarco et al., 1998; Ma et al., 1999). The influence of the differences in tissue composition on MC computation was investigated by du Plessis et al. (1998) for a set of 16 human tissues. They assessed the optimal number of tissue subsets to achieve 1% dose accuracy for millivolt photon beams, and concluded that seven basic tissue subsets were sufficient. By varying the physical mass density for lung and cortical bone, 57 different tissue types (excluding air) could be constructed, spanning a total CT range of 3000 HU. The division adopted by du Plessis et al. was 21 types of cortical bone ranging from 1100 to 3000 HU and 31 different lung tissues ranging from 20 to 950 HU. The remaining five types were chosen from the basic tissue subsets. The disadvantage of using the method proposed by du Plessis is that it does not use a stoichiometric conversion scheme and therefore is limited to the beam quality under consideration.

Stoichiometric conversion schemes were a large step forward in the accuracy of MC algorithms for patient dose calculations relative to the methods discussed previously. The stoichiometric conversion scheme from CT to electron (or proton stopping power) was initially proposed in Schneider et al. (1996) for proton MC dose algorithms. Initially, a set of materials of well-known atomic composition and density (usually plastic materials of varying composition and density to mimic human tissues) are CT scanned to measure the corresponding CT HU values. Next, the measured CT HU values are fitted by a theoretical equation interrelating the CT HU, mass density, atomic number Z, and atomic weight A of each material. Finally, the fitted parameters can be used to calculate the CT HU values for real tissues using tabulated composition data.

Schneider et al. (Schneider, 2000) proposed a CT conversion based on the stoichiometric conversion scheme, considering 71 human tissues and they adopted the total cross-section parameterization given by Rutherford et al. (1976). To reduce the effort of fitting the CT HU numbers to mass density and elemental weights, the authors created four sections on the CT scale within which the selected tissues were confined. Within each section, both the mass density and elemental weights of the tissues were interpolated.

8.4.1.1 Calculation of CT HU Number for Stoichiometric Conversion Schemes

In Rutherford et al. (1976), a parameterization of the cross section in the diagnostic x-ray energy range is presented:

$$\sigma(E) = Z_i K^{KN}(E) + Z_i^{2/3} K^{K^{\alpha\alpha}}(E) + Z_i^{4/3} K^{K^{\beta\beta}}(E),$$

(8.4)

where E is the energy, $\sigma(E)$ is the total cross section in units of barn/atom, $K^{KN}$ denotes Klein–Nishina coefficient, $K^{\alpha\alpha}$ is the Rayleigh or coherent scattering coefficient, and $K^{\beta\beta}$ is the
photoelectric coefficient. The linear attenuation coefficient, $\mu$, is now obtained by replacing the total cross section:

$$
\mu(E) = \rho N_A \sum_{i=1}^{n} \frac{w_i}{A_i} \sigma_i(E)
$$

$$
= \rho N_A \sum_{i=1}^{n} \left[ Z_i K_{K}^{KN}(E) + Z_i^{1.08} K_{\text{ac}}(E) + Z_i^{1.42} K_{\text{ph}}(E) \right] (8.5)
$$

where $\rho$ (g/cm$^3$) is the mass density, $N_A$ (mol$^{-1}$) is the Avogadro constant ($6.022045 \times 10^{23}$), $i$ is the element index, and $w_i$ is the element weight. Recall that the $\mu$ values are converted in HU using

$$
\text{HU} = \left( \frac{\mu}{\mu_{\text{H}_0}} \right) 1000 , \quad (8.6)
$$

where the CT HU is defined such that water has the value 0 and air has the value $-1000$. The equation then becomes

$$
\frac{\mu}{\mu_{\text{H}_0}} = \frac{\rho}{\rho_{\text{H}_0}} \left[ \sum_{i=1}^{n} \frac{w_i}{A_i} \left( Z_i + Z_i^{1.08} k_1 + Z_i^{1.42} k_2 \right) \right] (8.7)
$$

with $k_1 = K_{\text{ac}}/K_{\text{K}}$ and $k_2 = K_{\text{ph}}/K_{\text{K}}$, and the values of $(k_1, k_2)$ are dependent on the CT kilovolt values used for imaging and are determined experimentally by a least-square fit of the measured HU values to the previous equation, that is, minimizing the following expression:

$$
\sum_{i=1}^{n} \left( \frac{\mu}{\mu_{\text{H}_0}} \right) (k_i k_2) - \left( \frac{H_{\text{meas.}}}{1000} + 1 \right)^2. \quad (8.8)
$$

Additional predictions of the $k_1$ and $k_2$ values are also presented in Vanderstraeten et al. (2007) for a large group of CT scanners of different models and manufacturers for several European radiation oncology centers.

### 8.4.1.2 CT Hounsfield Units Interpolation Method

In this section, we present a brief description of the Schneider et al. method of interpolating within a tissue range, for tissue samples composed of only two components, where the components are denoted as $(\rho_1, w_{1,i}, H_1)$ and $(\rho_2, w_{2,i}, H_2)$ and the composite by $(\rho, w, H)$ and where $H_1$, $H_2$ represent CT HU values range and where $H_1 < H < H_2$. In addition, $w_i$ represents the proportion of the first component, while the second is obtained from $W_i = 1 - W_i$. A brief note on the meaning of $W_i$ and $W_2$ and how they differ from $w_{1,i}$ and $w_{2,i}$ while $w_{1,i}$ and $w_{2,i}$ represent the weighting within a tissue of the chemical elements such as oxygen, hydrogen, and so on, $W_i$ and $W_2$ represent the weighting of two reference tissues to form a new “interpolated” tissue such as combining osseous tissue and bone marrow to form different skeletal tissues. The new interpolated medium has the following weight per chemical element:

$$
w_i = W_i w_{1,i} + W_2 w_{2,i} = W_i (w_{1,i} - w_{2,i}) + w_{2,i} \quad (8.9)
$$

and mass density:

$$
\rho = \frac{m}{V} = \frac{m_{1}}{W_i (\rho_1 - \rho_i) + \rho_i} = \frac{\rho_2 \rho_i}{W_i (\rho_1 - \rho_i) + \rho_i} \quad (8.10)
$$

The linear attenuation coefficient of the new media can now be generated by substituting Equations 8.9 and 8.10 in Equation 8.5:

$$
\mu = \frac{\rho \rho_2}{W_i (\rho_2 - \rho_i) + \rho_i} N_A \left[ W_i \sum_i \left( \frac{w_{1,i} - w_{2,i}}{A_i} \sigma_i \right) \right] + \sum_i \left( \frac{w_{1,i} - \sigma_i}{A_i} \right) (8.11)
$$

The composite tissue with $H_i$, such that $H_1 < H < H_2$, is then given by

$$
W_i = \frac{\rho (H_2 - H)}{(\rho_2 H_2 - \rho_2 H_i) + (\rho_2 - \rho_i)H} \quad (8.13)
$$

The composite medium will then have the following mass density and elemental composition:

$$
\rho = \frac{\rho_2 H_2 - \rho_2 H_i + (\rho_2 - \rho_i)H}{H_2 - H_i} \quad (8.14a)
$$

$$
w_i = \frac{\rho (H_2 - H_i)}{(\rho_2 H_2 - \rho_2 H_i) + (\rho_2 - \rho_i)H} (w_{1,i} - w_{2,i}) + w_{2,i}, \quad H_1 \leq H \leq H_2 \quad (8.14b)
$$

### 8.4.1.3 Stoichiometric Conversion Method Based on Dose-Equivalent Tissue Subsets

In Vanderstraeten et al. (2007), the authors generalized the Schneider et al. (2000) method to generate subsets of tissue materials that were dosimetrically equivalent to within 1%. Their key assumption was that in the case of tissues that are situated close to the interface between two subsets, the dose differences
between them should not exceed 1%. In a previous study by Verhaegen and Devic (2005), the authors had showed the dosimetric importance of the CT calibration curve “tissue-binning” approach and the large dosimetric impact of tissue substitutes containing fluorine. Vanderstraeten et al. (2007) proposed the use of 14 dosimetrically equivalent tissue subsets (bins), where 10 were bone, to reduce significantly tissue-binning effects. The study was performed for a large number of different CT scanners obtained from nine different radiotherapy departments across Europe. These were then compared to the conventional five-bin scheme with only one bone bin. In Figure 8.1, the MC dose prediction for 5 and 14 bins of the CT is presented, where it is possible to see 5% differences at the interface.

The 10 bone tissue subsets used were generated to accurately model both the hydrogen (H) and calcium (Ca) content from 100 HU to beyond 1450 HU, consisting of cortical bone or denser bone. Both H and Ca contents strongly influence photon attenuation as pointed out in Seco and Evans (2006). In addition, the Ca content at the diagnostic energies will strongly increase the importance of the photoelectric and Rayleigh scattering terms because of their strong Z-dependence. The 14- and 5-bin approaches were then studied for a cohort of nine patients, where dosimetric differences of up to 5% were observed. However, DVH comparisons did not show any significant differences.

8.4.2 Dual-Energy X-Ray CT Imaging: Improved HU CT to Medium Conversions

Dual-energy x-ray CT imaging involves scanning an object with two significantly different tube voltages to obtain a better estimate of the effective atomic numbers, Z, and the relative electron densities, $\rho'$. In Torikoshi et al. (2003), the linear attenuation coefficient of a material at a set x-ray energy, $E$, is given by

$$\mu(E) = \rho'(Z^4 F(E,Z) + G(E,Z)),$$

where $\rho'$ is the electron density, $\rho'Z^4 F(E,Z)$ and $\rho'G(E,Z)$ are, respectively, the photoelectric and combined Rayleigh and Compton scattering terms of the linear attenuation coefficient. Both $F$ and $G$ are obtained by quadratic fits of the photoelectric and scattering terms of NIST attenuation coefficients (Berger et al., 2005).

A detailed method of performing dual-energy tissue segmentation is presented in Bazalova et al. (2008). Dual-energy CT material extraction was performed using the 100 and 140 kVp CT images with corresponding x-ray spectra. The authors found that the mean errors of extracting $\rho'$ and Z are 1.8% and 2.8%, respectively. Dose calculations were then performed using various photon beam energies of 250 kVp, 6 MV, and 18 MV, and 18 MeV electron beams and a solid water phantom with tissue-equivalent inserts. The dose calculation errors were particularly high for the 250 kVp beam, leading to 17% error due to a misassigned soft bone tissue-equivalent cylinder. In the case of the 18 MeV electron beam and 18 MV photon beams, they found 6% and 3% dose calculation errors due to misassignment. In the case of the 6 MV electron beam, dose differences were below 1%.

8.5 Deformable Image Registration

(A brief overview is given of deformable image registration; for more detailed description, please consult the Heath and Seco Chapter 7).
The use of deformable image registration varies from 4D radiotherapy to contour propagation, treatment adaptation, dosimetric evaluation, and 4D optimization (Keall et al., 2004; Rietzel, 2005; Trofimov et al., 2005; Brock et al., 2006). A variety of different approaches exist to the nonrigid registration; Wang et al. (2005) used an accelerated Demons algorithm, Yang (2007) employed B-spline image registration with normalized cross-correlation metric, and Rietzel and Chen (2006) used a B-spline method that respects discontinuity at the pleural interface. A variety of other methods exist, which use optical flow, thin-plate spline, calculus of variations, and finite element methods with different motion algorithms (Meyer et al., 1997; Guerrero et al., 2004; Lu et al., 2004).

8.5.1 Developing Dose Warping Approaches for 4D Monte Carlo

In Keall et al. (2004), the authors performed one of the first “pseudo” 4D MC dose calculations. The method used MC to calculate dose on each of the \( N = 8 \) individual 3D CT cubes representing an individual breathing phase. The dose distribution of each breathing phase was then mapped back to the end-inhale CT image set, using deformable image registration. The flowchart presented in Figure 8.2 represents the various steps of performing 4D MC dose calculation with deformable image registration. There are several stages to the process; from deforming contours of the anatomy from the reference phase to all CT datasets, to performing dose warping from the CT of the various breathing phases back to the reference dataset.

Deformable image registration has mostly been used in lung cancer dosimetric studies with MC (Flampouri et al., 2005; Rosu, 2005; Seco et al., 2008). A major issue with deformable registration algorithms is that no true one-to-one correspondence exists between the initial dataset (image or dose) and the registered dataset. This is a consequence of the initial voxels being rearranged (split or merged) to produce a new registered dataset. This occurs because registration algorithms either are statistical in nature or involve interpolation methods. Heath (2006) developed a method to circumvent errors introduced by deformable image registration mapping dose distributions. They developed the MC code defDOSXYZnrc, which uses a vector field to directly warp the dose grid on the fly while tracking individual particles. The major disadvantage of this method was the increased calculation time of the MC dose calculation by a factor of 2 or more.

In a more recent study by Siebers and Zhong (2008), a different approach was adopted to circumvent dose warping, while maintaining a fast MC dose calculation algorithm. The authors designated their method as the energy transfer method (ETM), where the separation between radiation transport and energy deposition is performed. The authors showed that some compensation mechanism must be provided that accounts for the rearrangement of the energy deposited per unit mass, particularly in the dose gradient regions. As an example of the ETM approach, consider two adjacent voxels \( A_1 \) and \( A_2 \) that merge into one voxel \( X_1 \), following image registration. The dose interpolation method calculates the dose, \( d(X_1) \), from \( d(A_1) \) and \( d(A_2) \) by

\[
d_{\text{INTERP}}(X_1) = \frac{d(A_1) + d(A_2)}{2} = \frac{(E(A_1) / m(A_1)) + (E(A_2) / m(A_2))}{2},
\]

where \( E(A_1), E(A_2), m(A_1), \) and \( m(A_2) \) are the energy deposited and masses of voxel \( A_1 \) and \( A_2 \), respectively. However, the authors point out that the correct interpretation of the dose should be

---

where the dose interpolation error is \( \varepsilon = |d^{\text{INTERP}} - d^{\text{ETM}}| \). The ETM approach allows energy mapping from multiple different source anatomies to multiple different "reference" anatomies simultaneously. In addition, the authors indicate that the method could also be useful in 4D IMRT optimization for 4D lung cases. In this case, beam intensities are simultaneously optimized on multiple breathing phases; thus, accurate dose evaluation is desired not only at a reference phase but also at each phase of the breathing cycle.

### 8.5.2 Comparing Patient Dose Calculation between 3D and 4D

Flampouri et al. (2005), studied the effect of respiratory motion on the delivered dose for lung IMRT plans, where a group of six patients were studied. For each patient, a free breathing (FB) helical CT and a 10-phase 4D CT scan was acquired. A commercial planning system was then used to generate an IMRT plan. The authors found that conventional planning was sufficient for patients with tumor motion less than 12 mm, where FB helical CT had small or no artifacts relative to the 4D CT dataset. The authors also found that conventional planning was not adequate for patients with larger tumor motion or severe CT artifacts. Their study indicated that CT reconstruction artifacts have far bigger effects than tumor motion. Therefore, a major conclusion was that for accurate 4D CT MC dosimetry, CT datasets with less artifacts and distortions are required. In addition, the authors also evaluated the minimum number of breathing phases required to recreate the 10-phase composite dose. They found that a three-phase composite dose would allow a 3% error relative to a 10-phase dose prediction, while five phases would achieve 0.5% error relative to the 10-phase dose. These results are in agreement with predictions by Rosu et al. (2007), where two phases (inhal and exhale) where shown to be sufficient to predict dose distribution in the respiratory motion. In Seco et al. (2008), the authors compared the dose distributions for 3D CT FB with 4D CT FB and 4D CT-gated treatments, to assess if gated treatments provided improved delivered dose distributions. The question addressed was: do we require gating to mitigate motion effects in delivered dose distributions? The breathing pattern of each individual patient was accounted for while studying interfraction and intrafraction effects of breathing in the delivered dose. The respiratory motion was recorded by the RPM system of Varian (Varian Medical Systems, Palo Alto, CA). The authors showed that the largest dosimetric differences occurred between inhal and the other breathing phases, including the FB scan.

To analyze the spatial differences between two MC dose distributions for different breathing phases as a function of the density, the authors defined a parameter called omega index, \( \Omega(\rho) \), or \( \Omega(\rho_f\text{VOI}) \), where the density, \( \rho \), is the independent variable, and \( f \) is a dose or dose difference distribution and VOI is the volume of interest for which the omega index is calculated:

\[
\Omega(\rho) = \int f(x) * \delta(\rho(x) - \rho) \, dx \over \int \delta(\rho(x) - \rho) \, dx,
\]

The parameter \( f \) can be either a 3D dose cube or a dose difference between two 3D cubes, that is, \( f = D_{MC_{\phi_1}} - D_{MC_{\phi_2}} \), where \( \phi_1 \) and \( \phi_2 \) are any breathing phases. The omega index indicates dose differences as a function of the density. In Figure 8.3, a comparison is given of the dose differences between inhale and FB dose calculations.

It is possible to observe dose differences of the order 3–5 Gy between inhale and FB that occur in both the low- and high-density regions of lung and bone, respectively. This indicates that 4D MC can improve not only accurate dose predictions within the primary tumor but also for the surrounding tissue, that is, lung or bone, which is considered part of the planning target volume (PTV) because of the expanding margin to account for motion and setup errors. The expanding area can encompass a large volume of both lung and bone, which may affect the final dose estimate to the PTV due to movement.

In addition, the authors also addressed the question of whether 4D-gated treatments are better than 4D free-breathing treatments in providing a better coverage of the PTV, where gated treatments have a far longer total delivery time and subsequent patient time on the couch. No significant dosimetric differences were observed between 4D CT FB and 4D CT-gated dose distributions; however, only a small group of three patients were assessed; so no generalization can be made. Larger differences (of approximately 3%) were observed between 3D FB and 4D FB, which could be attributed to various factors such as (i) differences in image reconstruction of serial versus helical CT, (ii) importance of the inhale phase in the final 3D FB CT, and (iii) coughing or breath-hold by the patient during the serial imaging process.

### 8.5.3 Intercomparison of Dose Warping Techniques for 4D Dose Distributions

In Heath et al. (2008), an intercomparison is performed of the various dose warping methods for 4D MC dose calculations in the lung. The methods compared were center-of-mass tracking, trilinear interpolation, and the defDOSXYZ method (Heath, 2006). No clinically significant dose differences were observed by the three methods. However, for the extreme case where motion was not accounted for in the treatment plan, the authors noticed an underestimate of the target volume dose by up to 16% by the remapping techniques. The authors also pointed out that the accuracy of the dose calculation is significantly affected by the continuity of the deformation fields from nonlinear image registration.
8.6 Inverse Planning with MC for Improved Patient Dose Calculations in Both 3D and 4D

Inverse planning for IMRT with any dose algorithm, including MC, involves the calculation of what is usually called the $D_{ij}$ matrix, which allows the conversion of a fluence map into a dose distribution by the following relationship:

$$ D_x(x) = \sum_{j \in \text{beamlet}} D_{ij} \chi_j, $$

where $D_{ij}$ is the respective dose influence and $\chi_j$ is a given fluence or beamlet intensity from a set of all beamlets. The $D_{ij}$ can then be calculated using pencil beam (PB), superposition-convolution, or MC dose algorithms. A major issue with the formulation given in Equation 8.16 is that usually delivery constraints and field size dependencies of output factors are not accounted for in the generation of $D_{ij}$ values. These have to be added after the planning stage, which can lead to suboptimal delivered dose distributions.

One of the initial attempts in using MC to perform 3D inverse planning was done by Jeraj and Keall (Jeraj and Keall, 1999) using the MCNP code. The authors performed optimization on MC-generated PBs, where an initial back-projection estimate was obtained with MCNP to generate an initial IMRT fluence. EGS4 was then used to calculate dose within the patient and therefore generate the required PBs for optimization of all beams, using a narrow-beam approach to generate each PB. The major issues with the method proposed by Jeraj and Keall (1999) was that each MC-generated PB did exactly model the spectrum exiting the linac, MLC delivery constraints were not modeled, and head scatter variations with field sizes and motion effects were not taken into account. More recently, Siebers et al. (Siebers and Kawrakow, 2007) proposed a hybrid method of 3D MC inverse...
Patient Dose Calculation

planning where an initial dose prediction was performed by PB algorithm. For subsequent iterations, the PB dose prediction was improved by using the MC algorithm such that the optimized plans were equivalent to an MC-based optimization. This produced a gain of 2.5 times over a full MC-based optimization.

Nohadani et al. (2009) proposed a 4D MC-based inverse planning approach using a broad beam approach, in contrast to the narrow beam approach of Jeraj et al. (1999). For each incident beam, the jaw setting does not change during IMRT delivery for a Varian linac, making it possible to generate an open-field phase space file just after the X/Y jaws. Therefore, the MC phase space after the jaws can be split into PB-like pixels from which particles can be sampled to produce a 3D PB dose distribution within the patient. In Figure 8.4a, we present a diagram showing how sampling a pixel in the PHase SPace (phsp) produces an MC–PB distribution within the patient.

The advantages of the broad beam approach for generating the MC–PB are: (i) it will account for heterogeneity effects in the patient IMRT calculation; (ii) if all PBs are combined into a homogeneous open field, the output factor will be that of standard open field, no further output corrections are required; and (iii) when delivering the IMRT profile with, for example, a multileaf collimator, the changes in output, due to field size changes during delivery, will be small and therefore will not affect the planned optimal dose distribution. In Figure 8.4b, the DVH comparing IMRT obtained with MC–PB versus standard PB is given. It shows that an IMRT plan based on PB may appear to be a better plan in terms of PTV coverage, hot spots, and organs at risk sparing. However, it deviates significantly from an MC-based optimised plan that closely mimics the reality of the deposited dose. The figure also illustrates that an otherwise optimal IMRT plan based on PB dose calculation may not correspond to the actual dose distribution because of these dosimetric errors. Therefore, errors in dose calculation cannot be neglected.

References


9

Monte Carlo Methods and Applications for Brachytherapy Dosimetry and Treatment Planning

9.1 Introduction

As in other branches of radiation therapy, Monte Carlo (MC) simulation has become an essential dosimetry tool in modern brachytherapy, playing key roles in both clinical practice and research. The most established application of MC methods in brachytherapy is the determination of dose rate distributions around individual radiation sources. Modern sources generally contain low energy radionuclides such as $^{198}$Au, $^{192}$Ir, or $^{131}$Cs (mean energies <0.05 MeV, referred to henceforth as low-energy sources) or higher energy radionuclides such as $^{192}$Ir and $^{137}$Cs (mean energies 0.355 and 0.662 MeV). There are also miniature x-ray sources generating at 50 kVp Bremsstrahlung spectrum. Both source geometries and clinical applications are quite variable. In low dose rate (LDR) brachytherapy, radioactive material and radio-opaque markers are encapsulated to form permanently implantable seeds. In high dose rate (HDR) brachytherapy using a remote afterloader, an iridium pellet is welded to the tip of a drive cable. Miniature x-ray sources with tungsten anodes also fall in the HDR category, even though they emit photons in the low-energy range. While inverse square law dependence is the dominating feature of brachytherapy dose distributions, photon attenuation and scatter build-up in the surrounding medium as well as radiation interactions within the source structure give rise to anisotropic dose distributions. The significant modulation of dose distributions must be properly modeled to attain clinically acceptable dosimetric accuracy, and these features are not readily derived from analytical methods such as the Sievert integral (Williamson, 1996). The complexities of experimentally measuring single-source dose distributions caused by the sharp dose gradients, low photon energies, and dose rates associated with brachytherapy make computational dosimetry techniques such as MC simulations an essential tool in brachytherapy (Williamson and Rivard, 2005). Furthermore, properly accounting for radiobiological effects in low-energy brachytherapy requires calculation of photon and electron spectra at various distances from sources, which is most accurately done by MC modeling.

The first computational efforts toward obtaining brachytherapy dose distributions are attributed to the 1960s work of Meisberger who derived 1D tissue-attenuation and scatter build-up factors for $^{198}$Au, $^{192}$Ir, $^{137}$Cs, $^{226}$Ra, and $^{60}$Co point sources (Meisberger et al., 1968), while Dale was the first to apply similar techniques to modern $^{125}$I in 1983 (Dale, 1983). Although MC modeling of a 3D brachytherapy source geometry was performed as early as 1971 by Krishnaswamy for $^{252}$Cf needles (Krishnaswamy, 1971), it took another decade for the field to fully embrace 3D modeling. Williamson showed in 1983, using 3D MC simulations, that the Sievert integral deviated by 5–100% from MC results for monoenergetic photons of energies lower than 0.3 MeV emitted from an encapsulated line source, emphasizing the need for accurate computational dosimetry.
Monte Carlo Techniques in Radiation Therapy

(Williamson et al., 1983). Burns and Raeside were the first to fully model a commercial $^{125}$I seed (model 6711, 3M, now GE HealthCare/Oncura), modeling the silver radiomarker, $^{125}$I distribution and titanium encapsulation, as shown in Figure 9.1, to obtain a 2D dose rate distribution (Burns and Raeside, 1988). Since the range in water of secondary electrons generated by 30 keV photons is less than 20 μm, Burns and Raeside did not transport electrons in their simulations and scored collision kerma using a track-length estimator (Williamson, 1987). Approximating absorbed dose by collision kerma is commonly employed by most MC codes used for brachytherapy dosimetry.

While estimating relative 2D dose distributions in medium around $^{125}$I sources from measurements and MC methods (Williamson and Quintero, 1988) was relatively common by the mid-1980s, the dose rate constant (dose at a reference point in medium per unit source strength) for low-energy seeds had not been definitively measured or calculated until the late 1980s. Williamson performed simulations in 1988 for models 6711, 6702, and 6701 3M sources as well as the National Institute of Standards and Technology (NIST) Ritz low-energy free-air chamber (Figure 9.2), used as air-kerma strength standard, to obtain dose rate constants (Williamson, 1988). By including the effects of 4.5 keV Ti K x-rays produced in the source encapsulation, first experimentally observed by Kubo (1985), Williamson showed that $^{125}$I absolute dose rates obtained from semiempirical methods were overestimated by 10–14%.

The 1990s saw several studies comparing MC simulations of brachytherapy sources and thermoluminescent dosimeter (TLD) measurements (Kirov et al., 1995; Valicenti et al., 1995). Williamson observed good agreement (1–5%) between simulations and measurements for $^{125}$I when accounting for the measurement phantom medium in the simulations (Williamson, 1991). Excellent agreement was observed for $^{192}$Ir (2–3%), when the shape and size of the measurement phantom were modeled, although the influence of its composition was found to be less important in this energy range (Williamson, 1991). In a significant series of articles, Williamson’s group performed extensive benchmarking of MC photon-transport calculations against precision diode measurements in water showing that MC accurately (1–3%) reproduced both relative and absolute dose rates across the entire brachytherapy energy range in both homogeneous and heterogeneous phantom geometries (Williamson, 1991; Williamson et al., 1993; Perera et al., 1994; Das et al., 1996, 1997a,b). These results confirmed that the MC methodology applied to brachytherapy dosimetry was mature and sufficiently accurate and robust to support clinical dosimetry. This was reinforced by the American Association of Physicists in Medicine (AAPM) Task Group No. 43 (TG-43) requirement that at least one experimental and one MC determination of dosimetry parameters be published before using a source clinically, essentially making MC simulations an industry standard (Rivard et al., 2004). Investigators tackled evermore complex problems, such as the modeling of multicoil radioactive stents used in the treatment of arterial restenosis (Reynaert et al., 1999), illustrated in Figure 9.3. However, MC results should not be trusted blindly, as illustrated by the significant dose-estimation errors resulting from the use of obsolete low-energy cross-section data (Demarco

---

**FIGURE 9.1** First 3D model of a 6711$^{125}$I source used in MC simulations by Burns and Raeside (1988). All dimensions in mm. The silver cylinder is 3 mm in length and 0.5 mm in diameter. (From Burns, G. S. and D. E. Raeside. 1988. Med Phys 15:56–60. With permission.)

Monte Carlo Methods and Applications for Brachytherapy Dosimetry and Treatment Planning

et al., 2002; Bohm et al., 2003; Reniers et al., 2004), a problem first pointed out by Williamson (1991). For brachytherapy dose distributions in the <50 keV energy range, where energy deposition is dominated by photoelectric absorption, even 1–2% errors in the photoelectric cross section can give rise to dose computation errors as large as 10–15% at 5 cm distance. This led to the adoption of modern cross-section libraries derived largely from theoretical quantum mechanical models (Cullen et al., 1997). The reader is referred to Williamson and Rivard for a more detailed discussion of this complex issue (Williamson and Rivard, 2009).

The rising popularity of prostate seed implantation in the United States, increasing from 5,000 in 1995 to about 50,000 in 2002, fuelled a rise in the number of commercially available brachytherapy seeds and designs (Grimm and Sylvester, 2004). While the initial 1995 TG-43 report presented consensus dosimetry parameters for one 103Pd and two 125I seeds (Nath et al., 1995), its 2004 low-energy seed update (TG-43U1) presented data for eight seed models (Rivard et al., 2004), while the 2007 supplement presented data from an additional eight seed models (Rivard et al., 2007). Yet another supplement is in preparation for the remaining commercially available low-energy photon-emitting sources. The recently published joint AAPM/European Society of Radiation Oncology (ESTRO) report (Perez-Calatayud et al., 2012) on high-energy brachytherapy dosimetry applied the AAPM prerequisite (Li et al., 2007) to 21 192Ir, 137Cs, and 60Co source models. This increasing proliferation of new brachytherapy sources and dosimetry datasets was associated with a rapidly growing number of MC-related brachytherapy publications in the peer-reviewed literature (Figure 9.4) and resulting in the 2001 “seed policy”

FIGURE 9.3 (a) EGS4 MC model of the ASC Multilink stent. (b) Scanning electron microscope image of the stent. (From Reynaert, N. et al. 1999. Med Phys 26:1484–1491. With permission.)
Monte Carlo Techniques in Radiation Therapy

of the Medical Physics journal, which limited TG-43 dosimetric parameter papers to technical notes and requested the authors to consider other publication venues for TG-43 seed parameter papers that did not exhibit sufficient scientific novelty to merit Medical Physics publication.

The use of MC methods in brachytherapy now goes beyond single-source dosimetry. An important area of current research is application of MC simulations to perform patient-specific dose calculations. This effort is necessary to overcome the limitations of TG-43-style table-based source-superposition algorithms, which include neglecting interseed attenuation (ISA) and tissue heterogeneities for low-energy seed implants (Burns and Raeside, 1989; Chibani et al., 2005; Carrier et al., 2007) and neglect of applicator shielding and partial scattering effects for higher-energy brachytherapy procedures (Valicenti et al., 1995; Gifford et al., 2005; Poon et al., 2006; Poon and Verhaegen, 2009). Several MC dose calculation platforms, generally based on computed tomography (CT) images, have been presented in the literature (Chibani and Williamson, 2005; Taylor et al., 2007; Afsharpour et al., 2012; Sampson et al., 2012). For low-energy sources, the most important challenge to accurate patient-specific dosimetry is accurate voxel-by-voxel assignment of photon cross-section tables.

In this chapter, we will cover reference quality and clinical dose calculation methods for single-source dosimetry, clinical patient-specific treatment planning, and thoughts on future technological advancements in brachytherapy.

9.2 Single-Source Dosimetry

The established application of MC simulations is determination of the single-source dose distribution for calculation of patient dose distributions using table-lookup methods. In addition to describing the general process for source dosimetric characterization, this section will highlight issues unique to MC and other Boltzmann equation numerical solutions, including influence of seed design, phantom size and composition, and other effects not normally encountered in applying experimental and semiempirical dosimetry methods.

9.2.1 TG-43 Dosimetry Parameter Reference Datasets

Current brachytherapy treatment planning systems (TPS) utilize the dose calculation formalism of the AAPM TG-43 report (Nath et al., 1995; Rivard et al., 2004, 2007). The TG-43 report assumes that a brachytherapy source produces a cylindrically symmetric dose distribution and permits the treatment planner to determine patient dose distributions about the source via table lookup in a polar coordinate system using nonlinear interpolation.

When the source orientation is known, the 2D formalism of the TG-43 report can be used. In this formalism, contribution of each source to a point of interest (POI) is a function of polar angle $\theta$ and distance $r$ from the source center to the POI (Figure 9.5). Sources in which the orientation is known include HDR $^{192}$Ir or $^{60}$Co sources (photon-emitting brachytherapy sources with average energy $>50$ keV, referred to henceforth as high-energy sources), that is, those delivered with a remote afterloading unit (RAU), LDR $^{137}$Cs or $^{192}$Ir sources, or low-energy LDR sources such as $^{125}$I, $^{103}$Pd, and $^{131}$Cs when the imaging modality (typically CT or three-film projection techniques (Pokhrel et al., 2011) permits adequate resolution to discern the capsule orientation. When clinical imaging is inadequate to discern the capsule orientation, the 1D formalism of the TG-43 report is used, in which the polar angle dependence of single-seed contributions is ignored (Corbett et al., 2001; Lindsay et al., 2001).

#### FIGURE 9.5

Coordinate system employed by the TG-43 formalism for a source with active length $L$. The POI is indicated by $P(r, \theta)$. (From Nath, R. et al. 1995. Med Phys 22:209–234. With permission.)
9.2.1.1 Dose Calculation Formalisms

The 2D dose calculation formalism of the AAPM TG-43 report is given as

\[
\tilde{D}(r, \theta) = S_K \cdot \Lambda \cdot \frac{G_L(r, \theta)}{G_L(r_0, \theta_0)} \cdot g_\perp(r) \cdot F(r, \theta)
\]  

(9.1)

where the dose rate as a function of \( r \) and \( \theta \) (see Figure 9.5) is equal to the product of the air-kerma strength \( S_K \) (a measure of the brachytherapy source strength (Rivard et al., 2004)); the dose rate constant \( \Lambda \) (defined as the ratio of the reference dose rate \( D(r_0, \theta_0) \), where \( r_0 = 1 \) cm and \( \theta_0 = 90^\circ \) and \( S_K \) as \( \Lambda = D(r_0, \theta_0)/S_K \)); the geometry function \( G_r(r, \theta) \) at any point \( r \) and \( \theta \) using the line-source (for a source of active length \( L \)) approximation divided by its value \( G_r(r_0, \theta_0) \) at \( r_0 \) and \( \theta_0 \); the radial dose function \( g_\perp(r) \) based on the line-source approximation; and the 2D anisotropy function \( F(r, \theta) \).

The 1D dose calculation formalism of the AAPM TG-43 report is given as

\[
\tilde{D}(r) = S_K \cdot \Lambda \cdot \frac{G_L(r, \theta)}{G_L(r_0, \theta_0)} \cdot g_\perp(r) \cdot \phi_{m}(r)
\]  

(9.2)

where the dose rate as a function of \( r \) only is equal to the product of \( S_K \), \( \Lambda \), the geometry function \( G_L(r_0, \theta_0) \) at any distance \( r \) using the line-source approximation divided by its value \( G_L(r_0, \theta_0) \) at \( r_0 \) and \( \theta_0 \); the radial dose function \( g_\perp(r) \) using the line-source approximation, and the 1D anisotropy function \( \phi_{m}(r) \).

These dosimetry parameters for the 2D and 1D dose calculation formalisms are described in great detail within the 2004 AAPM update to the TG-43 report (TG-43U1) (Rivard et al., 2004). Of note is that all parameters take a value of unity at the reference position \( (r_0, \theta_0) \) except \( S_K \) and \( \Lambda \) such that \( \tilde{D}(r_0, \theta_0) = S_K \cdot \Lambda \) and \( D(r_0) = S_K \cdot \Lambda \).

Using equations and recommendations covered in TG-43U1, the MC dosimetry investigator extracts tables of TG-43 parameters from the measured or MC-computed dose distribution, which is typically performed using detectors arranged in a sparse polar coordinate grid. In turn, these parameters are then imported into a clinical brachytherapy TPS using the same formalism. Recovering the dose distribution in the original MC grid for each source modeled in the TPS is an essential institution-specific quality assurance test (Rivard et al., 2004).

9.2.1.2 MC Dose Calculation Method

The process for calculating dose rate distributions in the vicinity of a brachytherapy source has been described in the joint TG-138 report by the AAPM and the ESTRO on brachytherapy dosimetric uncertainties (DeWerd et al., 2011), the revised TG-43 report, and a recent review article (Williamson and Rivard, 2009). In short, the dosimetry parameters calculated in liquid water \( \Delta D(r, \theta) \) (where \( \Delta D \) means absorbed dose in the inherent units of the MC output, usually Gy/simulated primary photon or a multiple thereof) are determined separately from \( \Lambda \). We describe a typical MC dose simulation process as a roughly chronological sequence.

9.2.1.2.1 Brachytherapy Source Design

Before initiating MC simulations of brachytherapy dose rate distributions, the investigator must have an accurate and complete geometric model of the seed. This includes as many as possible of the following: the assembly process, the atomic composition, density, shape, dimensions, and location of each component along with tolerances in any of these features. In addition, the geometry and location of the radionuclide distribution, along with any radioactive contaminants or nonradioactive filler material, must be known. Often, this distribution is a very thin layer the thickness of which can only be known approximately. Typically, the MC investigator starts with manufacturer-provided construction drawings of an average source possibly along with proprietary descriptions of the chemical fabrication process, from which additional descriptive details can be inferred. TG-43U1 recommends that dosimetry consultants independently validate these specifications. As described elsewhere (Williamson and Rivard, 2009), validation techniques include direct mechanical measurement of nonradioactive seeds and preassembly internal components; pin hole autoradiography; transmission radiography; and low-magnification optical or electron microscopy of longitudinally sliced-open seeds or individual seed components. Because the MC simulation process tries to estimate the physical dose rate distribution, any differences between the simulated and real geometry may result in erroneous results. While some simulation errors are more forgiving than others (such as for thin-walled, high-energy photon-emitting brachytherapy sources), it is not always clear what is the quantitative sensitivity of a given design feature to the resultant dose rate distribution. TG-43U1 and TG-138 recommend that dosimetry investigators perform careful uncertainty analysis by performing MC simulations over an ensemble of source geometries sampled from the estimated tolerances of geometric parameters whose mean values cannot be accurately determined, as illustrated by a recent study of the Model 6711 129I seed (Dolan et al., 2006).

Some sources have internal components that move under the influence of the gravitational forces, resulting in an orientation-dependent geometry not evident upon external examination of the source (Rivard, 2001, 2007; Williamson, 2002; Rivard et al., 2004). This is most prevalent for low-energy LDR sources where the internal components are not rigidly attached to the capsule and the radiation emissions are highly attenuated by radio-opaque markers. This is illustrated in Figure 9.6 for \(^{203}\text{Pd}\) sources. For these sources, the dose rate can vary by more than a factor of two within a few millimeters of the source. In this circumstance, MC simulation is a powerful tool for performing sensitivity analyses to understand how the dose rate distribution is modulated by repositioning of internal source components. Unlike LDR sources, HDR sources have their capsule attached to a drive cable for translating the source through catheter(s) in a patient. Because the source design may not be rigid, and since the HDR RAU drive cable is flexible, it may not be clear...
which source geometry best represents the most probable dose distribution. Simulations of ensemble source configurations are needed to address the impact of dynamic source geometries and to develop practical solutions (Mikell and Mourtada, 2010).

9.2.1.2.2 Brachytherapy Source Radiation Emissions

In general, brachytherapy sources utilize radionuclides as the radiation source; electronic brachytherapy sources are an exception. The nuclear disintegration process is well understood with photon energies, emission rates, and branching ratios often known to three or more significant digits for common radionuclides.

Brachytherapy sources are designed around a single radionuclide. However, radioactivity activation and fabrication processes often result in potentially significant concentrations of contaminant radionuclides even after efforts to improve radio-purity via downstream radiochemistry. Since radiopurification processes are not perfect, all brachytherapy sources contain radioimpurities. While the dosimetric influence of radioimpurities has not been sufficiently examined in the literature, the joint AAPM/ESTRO TG-167 report conservatively recommends that radioimpurities be minimized to such levels that their dosimetric contributions over the range of clinically relevant distances in the vicinity of the implant should be ≤5% of the dosimetric contributions of the primary radionuclide. In addition, the working life of the source must be considered for radioimpurities that are longer-lived than the intended radionuclide. The MC dosimetry investigator is advised to consider the impact of radioimpurities on the recommended dose distribution.

In addition to radioimpurities, all relevant radiation emissions for a given radionuclide should be included in the simulation, where “relevant” means affecting the dose distribution over the range of clinical interest by more than 2–3%. While neutrino emissions are dosimetrically irrelevant, many common radionuclides disintegrate via beta-decay with subsequent x-ray and gamma-ray emissions. Not all electrons (or positrons) generated within the capsule interior are fully attenuated by the source structure and may deposit dose (or bremsstrahlung x-rays) exterior to the capsule (Granero et al., 2011). Electronic brachytherapy sources generate characteristic x-rays in addition to bremsstrahlung photons in various internal components, which may also contribute to clinically significant energy deposition outside the source capsule (Liu et al., 2008). At a minimum, MC codes used in low-energy dosimetry should accurately model emission of K- and L-shell characteristic x-rays following ejection of inner-shell electrons via knock-on electron collisions, photoelectric effect, or inelastic photon scattering. The MC dosimetry investigator is advised to perform calculations to benchmark the code preceding simulation of a clinical brachytherapy source. This is discussed in greater detail in Section 9.2.1.2.5.

9.2.1.2.3 Simulating Phantoms for Reference Quality
or Experimental Dosimetry Measurements

After source radiation is generated within the capsule, it may pass from the source into the surrounding phantom. A liquid water sphere is the standard phantom for presenting reference quality, single-source brachytherapy dose distributions. For low-energy sources, the sphere radius is 15 cm with the brachytherapy source being studied located at the center of this sphere (Rivard et al., 2004). For high-energy photon-emitting sources (>50 keV photons), the recommended reference phantom radius is 40 cm (Perez-Calatayud et al., 2012). Given the ranges of back-scattered photons, the phantoms permit acquisition of dose rate data within 1% of the unbounded phantom dose distribution (Perez-Calatayud et al., 2004; Melhus and Rivard, 2006) for distances up to 10 and 20 cm, respectively, for low- and high-energy sources. In practice, simulating a larger phantom produces dose distributions at the phantom periphery quantitatively closer to that of an infinitely large phantom, with a penalty of longer simulation times for the same number of particle histories. Regardless, the MC dosimetry investigator should state the used phantom dimensions and employ actual-to-reference phantom corrections factors, so that the final recommended dose distribution adheres to the relevant AAPM recommendation.

The AAPM TG-43 dose calculation formalism ignores patient tissue composition and is based upon specification of absorbed dose and transport of radiation in liquid water. The AAPM TG-43U1 report (Rivard et al., 2004) specifies liquid water to consist of exactly two parts hydrogen and one part oxygen with a mass density of 0.998 g/cm³ at a temperature of 22°C. For low-energy brachytherapy dosimetry, cross sections for coherent scattering are very sensitive to intermolecular forces and molecular bonds. Hence, the atomic mixture rule, which generally supports accurate synthesis of cross sections for compounds and
mixtures, is not applicable to coherent scattering. To accurately model coherent scattering, first, calculated $\sigma_{\text{coll}}$ values should be recalculated using liquid or molecular water atomic form factors (Morin, 1982) and the water coherent scattering cross sections extracted from NIST XCOM or EPDL97 libraries discarded, as they are based upon the mixture rule. Additionally, dosimetry investigators utilizing experimental techniques need to be aware that the phantoms used for dose measurement often deviate significantly from liquid water (Williamson, 1991; Patel et al., 2001). As recommended by TG-43U1, brachytherapy experimentalists are expected to use MC simulation to calculate reference-to-experimental phantom corrections factors so that the final experimental results represent absorbed dose to liquid water in the spherical liquid water reference phantom. Finally, the reader must remember that the patient tissue compositions, densities, and dimensions deviate significantly from the reference geometry, which can result in large discrepancies between estimated and delivered dose, especially for low-energy sources.

The AAPM TG-186 is the first concerted attempt to formulate guidelines for assigning cross sections and densities, other than unit-density water, to organs and tissues based upon patient-specific imaging (Beaulieu et al., 2012).

9.2.1.2.4 MC Radiation Transport Codes

Radiation transport following radioactive decay is a complex phenomenon, consisting of attenuation and scattering of photons, as well as transport of beta rays and secondary electrons via elastic collisions with nuclei; elastic and inelastic collisions with orbital electrons; and bremsstrahlung production. Because electrons experience on the order of $10^6$ discrete collisions as they slow down, resulting in impractically long computing times, electron transport, if included at all, is modeled via the approximate “condensed history” approach (Berger, 1963; Kawrakow and Bielajew, 1998). In this approach, a much smaller number of randomly sampled condensed history steps is simulated, each of which represents the collective effects of many discrete collisions by means of multiple scattering distributions and the continuous slowing down approximation (CSDA), with either stochastic or analytical corrections to account for random fluctuations in energy loss about the CSDA energy loss. Despite these approximations, condensed history transport has been demonstrated to accurately characterize absorbed dose distributions over a wide range of circumstances with reasonable computing times.

Photon interactions may also be simplified. An MC code may simulate K-edge characteristic x-rays, but not consider L-edge x-rays for low-Z elements, or the atomic relaxation process may be simplified, including only the most probable K- and L-shell energy transitions for high-Z elements. Further, radiological interactions at energies less than a few keV will be subject to molecular binding effects, which are not accounted for in most current MC codes in use, with the possible exception of coherent scattering as described above. Even the impulse approximation, which models the influence of bound orbital electrons on Compton and Thomson scattering via the form factor approach, fails to model the influence of orbital electron momentum distributions, which give rise to a spectrum of scattered photon energies centered about the Compton scattered photon energy, resulting in Compton Doppler broadening (Ribberfors and Carlsson, 1985). However, even in the low photon energy range, the dosimetric differences between the form factor and consistently implemented free-electron scattering models are small (Taylor and Rogers, 2008a; Williamson and Rivard, 2009).

Neglecting electron transport altogether, and simulating only photon transport, substantially reduces the computing burden in brachytherapy. This simplification is exploited by most dosimetry investigators and all specialized codes intended for treatment planning applications. It approximates absorbed dose by collision kerma, which in turn assumes equivalence of secondary electrons. This approximation is valid everywhere for low-energy brachytherapy sources, where electron ranges are of the order of 10 $\mu$m. However, for higher-energy $^{192}$Ir and $^{137}$Cs sources, which have secondary electron and beta-ray ranges of 1–3 mm, charged-particle equilibrium approximation may introduce significant errors near metal–tissue interfaces and near sources. For example, dose errors exceeding 15% at distances less than 1 mm from an HDR $^{192}$Ir source have been observed (Taylor and Rogers, 2008b; Ballester et al., 2009). In addition, hot spots due to beta-ray leakage near the end of LDR $^{192}$Ir seeds have been observed experimentally (Chiu-Tsao et al., 2004). High-energy photon-emitting sources such as $^{192}$Ir and $^{137}$Cs may not be accurately simulated at distances within a few millimeters (DeWerd et al., 2011). While one must be aware that MC photon transport codes give rise to >5% errors at 1–2 mm distances from $^{192}$Ir and $^{137}$Cs sources (Perez-Calatayud et al., 2012), photon-only transport solutions produce acceptable results for conventional interstitial and intracavitary applications where the clinically relevant distance range is 3–50 mm. For applications where near-zone dosimetry is important, for example, $^{192}$Ir-based intravascular brachytherapy or rigorous estimation of mucosal doses in contact with metal applicators, coupled photon–electron calculations and inclusion of the radionuclide beta-ray spectrum may be necessary.

9.2.1.2.5 Photon Cross-Section Libraries and Monte Carlo Scoring Functions

While both low- and high-energy source dose distributions are dominated by $1/r^2$ falloff, radiation interactions within the source capsule and in the surrounding medium significantly modulate the dose distribution, requiring MC simulation or other Boltzmann equation solutions to predict doses with clinically acceptable accuracies. Having finalized the source and phantom geometries and having chosen the collisional physics model, appropriate cross-section tables must be selected and assigned to each material in the simulation geometry. In addition, the detector geometries and estimator, that is, the mathematical function that estimates the absorbed dose contribution to each detector from each simulated history, must be selected.

Photon interaction cross sections are compiled into data libraries and are generally based on quantum mechanical models.
of each scattering and absorption process based upon approximate models of orbital electron wave functions, which have previously been validated by comparison to available experimental measurements (Hubbell, 1999). These libraries are maintained by organizations such as the National Nuclear Data Center of Brookhaven National Laboratory in New York, the Nuclear Data Section of the International Atomic Energy Agency in Vienna, and the Nuclear Energy Agency Data Bank of the Organization for Economic Cooperation and Development in Paris. However, the main clearinghouse for evaluated cross-section libraries and public domain transport codes is the Radiation Safety Information Computational Center (RSICC) (http://www-rsicc.ornl.gov), which is located at Oak Ridge National Laboratory. For modest fees, cross-section libraries and many codes for cross-section preprocessing and MC simulation can be obtained. Cross-section libraries typically consist of tables of partial cross sections for photoelectric effect, coherent scattering, incoherent scattering, and pair production for each element on a coarse logarithmic energy grid. In addition, atomic form factors and incoherent scattering factors for each element are available. The most modern libraries, for example, EPDL-97, have subshell-specific photoelectric cross sections and form factors and extensive tables of orbital electron transition and fluorescent yields so that atomic relaxation processes (the cascade of characteristic x-rays and Auger- and Coster-König electrons emitted following ejection of an inner shell electron by one of the photon or electron interaction processes) can be properly simulated. For brachytherapy, the most important issue is to select a library that contains up-to-date (post 1983) photoelectric effect and scattering cross sections. Fortunately, all of the modern libraries, for example, EPDL97 from Lawrence Livermore Laboratory, DLC-146 from RSICC, and XCOM from NIST, are all based on the same theoretical models, despite having many differences in format and extensiveness of compiled data. Specialized libraries of up-to-date cross sections in EGSnrc, Geant, and MCNP formats are also available. A more detailed review of both modern and obsolete cross-section libraries and formats is given elsewhere (Williamson and Rivard, 2009). These cross-section libraries can subtend several hundred megabytes and have complex numerical formats, for example, ENDF-B. These sparse tables therein are intended for interpolation onto finer grids using specific nonlinear interpolation schemes, most commonly log-log-linear interpolation for photon data. Most libraries come equipped with programs designed to access the database and preprocess the data into user-specified formats. When using the library data, the MC investigator must be careful to employ appropriate interpolation software so as to avoid large errors in the low-energy range (Demarco et al., 2002). Another related quantity that may be needed, depending on the estimator selected, is the mass-energy absorption coefficient. Tables of mass-energy absorption coefficients, consistent with the NIST XCOM or EPDL97 cross sections may be obtained from the NIST website (http://www.nist.gov/pml/data/xraycoef/index.cfm), in which elements $Z < 92$ are tabulated over a wide energy range. Tables for user-specified mixtures and compounds may also be obtained using the mixture rule. It is essential that the data used for scoring be derived from the same cross-section data used to transport the photons. For example, if the free-electron scattering model (no coherent scattering or electron-binding corrections to Klein–Nishina scattering) is used for transporting photons, then the NIST mass–energy absorption coefficients must be recalculate by replacing the term associated with incoherent scattering energy deposition with its Klein–Nishina counterpart.

Calculation of absorbed dose or collisional kerma is then a function of the radiation energy and medium (e.g., water). Uncertainties in the phantom composition or its radiological cross-section library will become more pronounced as distance from the source increases and systematic errors accumulate.

The final choices to be made are selection of a scoring grid and an estimator. More detailed treatments of this complex subject can be found elsewhere (Williamson, 1987; Williamson and Rivard, 2009). The simplest choice is the analog estimator, in which only those simulated collisions that occur within the detector volume contribute to dose. Invoking charged particle equilibrium (CPE), the dose contribution is computed as the difference between energy entering and leaving the detector divided by mass. While simple, the analog estimator is extremely inefficient. By using a variant of the track-length estimator (Williamson, 1987), efficiency can be improved 20- to 50-fold (Hedtjarn et al., 2002). The track-length estimator scores dose as $(\Delta l \cdot E/\Delta V) \cdot (\mu_w / \rho)$, where $\Delta l$ is the distance within the detector of volume $\Delta V$ traversed by incoming photon with energy $E$. Efficiency is improved since every voxel intersected by a photon flight path produces a nonzero dose score, greatly increasing the information that can be extracted from a finite sample of histories.

Since the analog and track-length estimators converge to the same value, collision kerma integrated over the detector volume, the detector grid must be selected carefully to minimize volume-averaging artifacts and to maintain acceptable statistics out to distances of 5–7 cm for low-energy brachytherapy. For single-source dose estimation, typically a spherical grid is used, with very thin detector elements near the source and thicker ones far from the source so as to improve statistics. By ignoring azimuthal angle, the detectors become a set of spherical segment shells that can vastly improve efficiency by exploiting cylindrical symmetry of the source. Since the TG-43 protocol requires specification of dose at geometric points, the MC investigator must correct each MC estimate for volume averaging, typically by using simple inverse square law-based correction factors (Ballester et al., 1997). An alternative to volume detectors is to use more complex dose-at-a-point estimators (Williamson, 1987; Li et al., 1993), in which analytic formulas are used to estimate the contribution from each simulated collision to dose at a geometric point. When using any kind of estimator, the statistical precision of the dose estimates, quantified in terms of standard deviation about the mean (67% confidence interval) should be carefully tracked as a function of distance. For volume detector estimators, detector size is always a trade-off between statistical precision (larger is always better) and spatial resolution, that is, reduced volume-averaging artifact, for which smaller is always better.
MC estimation of brachytherapy dosimetry parameters such as \(g(r), F(r, \theta), \) and \(\phi_{\text{tot}}(r)\) is performed using the selected detector grid and estimator by obtaining estimates of \(\Delta K(r, \theta) \) \(= \Delta D(r, \theta)\) (dose per simulated history) over the detector grid \((r, \theta)\) embedded in the reference liquid water sphere. By judicious choice of detector volumes and estimators, it should be possible to limit statistical uncertainties to <1% at distances less than 5 cm. Once volume-averaging corrections have been applied, the TG-43 relative parameters can be estimated using the equations given in the TG-43U1 report.

### 9.2.1.2.6 Dose Rate Constant

To estimate the value for \(\Lambda\), the ratio of \(\Delta D(r_1, \theta_1)\) to \(\Delta S_K\) (the air-kerma strength per simulated history) is obtained. Because MC methods generally estimate absorbed dose or collisional kerma in terms of energy imparted per number of particle histories performed or radionuclide disintegrations simulated, the results from the liquid water simulations need to be combined with simulations of the air-kerma strength in a suitable free-air geometry (vacuum or an air sphere). This way, the inherent normalization used by the MC code is irrelevant since both \(\Delta D(r_1, \theta_1)\) and \(\Delta S_K\) are expressed in terms of the same clinically irrelevant normalization. However, the MC normalization constant must be known when absolute dose distribution results are needed in terms of units of air-kerma strength. This process is described by Melhus and Rivard (2006) and Williamson and Rivard (2009) for calculating dose rate distributions per unit source strength \(U\), requiring knowledge of the number of photons emitted per disintegration. Thus, to compute dose around an actual brachytherapy source, the clinical medical physicist need only measure \(S_K\) (in a traceably calibrated reentrant well-type ionization chamber) and multiply this number by the MC calculated dose rate constant. Since the other TG-43 brachytherapy dosimetry parameters are ratios of doses computed in a single simulation run, it is crucial that the same source design and other starting conditions be used between the liquid water and vacuum (or air) simulations or else a scalar offset (percent-age systematic dose error) will be present throughout the entire range of calculated dose rates.

Vacuum is easy to model. Air is more complicated, requiring a mass density \((1.196 \text{ mg/cm}^3)\) at a temperature of \(22^\circ\text{C}\) for dry air at 101.325 kPa with mass compositions of 1.24%, 75.5268%, 23.1781%, and 1.2827% for C, N, O, and Ar, respectively (Rivard et al., 2004). As for the liquid water description in Section 9.2.1.2.4, this description for dry air is for standardization purposes since brachytherapy source strength (air-kerma strength) is calibrated for these conditions. There are several subtleties MC investigators must appreciate, which have been reviewed by Monroe and Williamson (2002). First, if air is employed, photon attenuation and scatter-build-up effects must be corrected in order to satisfy the AAPM's definition of \(S_K\), which is specified in free space. Second, care must be taken to filter out or suppress any low-energy contaminant photons below 5 keV, as recommended by the TG-43U1. Especially if a vacuum-filled calibration range is simulated, failure to suppress Ti characteristic x-rays could inflate \(\Delta S_K\) by 20%. Finally, care should be taken to simulate any known departures of the NIST primary standard from the \(S_K\). This was first demonstrated by Williamson in 1988 (Williamson, 1988), who found it necessary to simulate the Ritz free-air chamber and the NIST procedure (Loftus, 1984) for air-attenuation correction in order to correctly model the impact of Ti capsule characteristic x-ray generation on the comparison of theoretical and measured dose rate constants. More recently, it was found that the sharp edges of the palladium-plated carbon pellets of the model 200\(^{103}\)Pd seed induced significant and distance-dependent anisotropy of the polar fluence profile near the transverse axis (Monroe and Williamson, 2002). This necessitated a detailed simulation of the wide-angle free-air chamber (Seltzer et al., 2003), which integrated over this region of polar anisotropy, significantly deviating from the definition of air-kerma strength and altering the dose rate constant by as much as 12%.

### 9.2.2 Calculation of Spectral Effects and Relative Biological Effectiveness

#### 9.2.2.1 Spectral Effects and Detector Response

MC simulations are also excellent sources of information on photon and electron spectra at various distances from brachytherapy sources. These data are needed to correct the readings of dosimetry detectors that may have an energy-dependent response (film, TLD, ionization chamber, etc.). This is essential for both accurate absolute and relative dosimetry near brachytherapy sources. Das et al. (1996) used the Monte Carlo photon transport (MCPT) (Williamson and Meigooni, 1995) code to investigate the energy response of diodes and two sizes of TLDs to a range of photon energies from superficial x-rays to \(^{192}\)Ir and \(^{60}\)Co sources. They found a constant ratio between measured detector signal and MC calculated dose to the detector-sensitive volume for all detectors used, independent of photon energy. This indicates that MC simulations are indeed able to model energy response of detectors in the brachytherapy energy range. The absolute response function of the detectors used was also found to be proportional to absorbed dose by the active volume of the detector. However, other recent investigators (Davis et al., 2003; Nunn et al., 2008) claim that the relative intrinsic TLD-100 response at \(^{125}\)I energies (reading per unit dose to the active detector volume at \(^{125}\)I relative to that at 4 MV x-rays or \(^{60}\)Co) deviate from unity by 5–10% due to the impact of higher LET low-energy photons on the process of TLD signal formation. On the other hand, comparison of MC doses to TLD measurements for many published data sets suggests that a relative energy response factor closer to that computed by MC alone is more accurate (Williamson and Rivard, 2009). The appropriate choice of TLD-100 relative energy response correction remains unresolved at the moment.

A recent example of the utility of MC simulations to determine energy response with respect to \(^{60}\)Co radiation of various detector materials embedded in various phantom materials employed in brachytherapy is the study by Selvam and Keshavkumar (2010). They used EGSnrc codes and a large cavity approach
to calculate the energy response of $^{125}$I and $^{169}$Yb sources as a function of distance and detector material thickness, but they did not model detailed detector geometries. Their work (and references therein) clearly shows that energy correction factors need to be taken into account in brachytherapy and that MC simulation offers a convenient and accurate way to obtain them. Failure to account for these energy response effects may lead to severe over/underestimation of measured dose particularly in low-energy brachytherapy. In another recent application using MOSFET detectors, the mean photon energy from an $^{192}$Ir source was found to decrease significantly with distance from the source by more than 100 keV over about 10 cm, leading to an increase in dose response by 60% (Reniers et al., 2012).

Another powerful application of MC simulations in brachytherapy is the determination of dosimetry detector perturbations in the—sometimes very close—vicinity of brachytherapy sources. Reynaert et al. (1998) modeled an NE2571 chamber in detail, which was used to perform dosimetry close to a pulsed dose rate (PDR) $^{192}$Ir source in a water phantom. MC simulation is often the only method to reliably determine these perturbation factors, as discussed in other chapters in this volume. The use of MC-based correction factors for TLD dosimetry in brachytherapy has been reviewed in detail by Williamson and Rivard (2009).

### 9.2.2.2 Radiobiological Effectiveness

Knowledge of photon and electron spectra is also needed for applying biological damage models. It has been known for years from radiobiology that photons with energies below about 50 keV are more potent in inducing DNA damage than higher-energy photons (Hill, 2004). This finding, however, is rarely taken into account in clinical brachytherapy practice. As an example we discuss here an EGSnrc MC model of an electronic brachytherapy source, operated at 50 kVp. Reniers et al. (2008) simulated the initial yield of single- and double-strand DNA breaks (SSB and DSB), which allowed estimation of the relative biological effectiveness (RBE). The MC model was first used to generate x-ray spectra at various distances from the source and to calculate electron spectra from the photon spectra in several media (Figure 9.7). These data were then input into an empirical “Monte Carlo damage simulation” program derived from track structure considerations (Semenenko and Stewart, 2004) to estimate the initial yield of single- and double-strand breaks in DNA. From this, the RBE was derived.

The results indicate a substantially increased DSB yield for the electronic brachytherapy source compared to $^{60}$Co or $^{192}$Ir reference radiations, leading to an enhanced RBE for DSB by 40–50%. This RBE increase was mostly caused by Compton and Auger electrons with energies below 8 keV (Figure 9.7). The RBE estimate for the low-energy x-ray source was found to be very similar to the measured and calculated RBE for the low-energy gamma-ray brachytherapy radionuclide $^{125}$I (Reniers et al., 2004). These findings should be taken into account if the HDR electronic brachytherapy source is intended to replace brachytherapy with the commonly used HDR $^{192}$Ir radionuclide.

### 9.3 Treatment Planning

#### 9.3.1 Introduction

The TG-43 formalism, based on the work of the Interstitial Collaborative Working Group (ICWG) (Anderson et al., 1990), was a major step forward in brachytherapy dose calculation. It replaced the use of semiempirical calculation methods based on apparent activity, equivalent mass of radium, exposure rate constants, and tissue-attenuation coefficients, quantities that are not source model but only radionuclide dependent. By employing dosimetry parameters that are dependent on the radionuclide, its distribution in the source, and the source geometry itself, and by recommending measured or calculated consensus datasets, the formalism of the TG-43 protocol, and its subsequent updates, improved standardization and accuracy in brachytherapy dose calculations. Fifteen years later, the widely cited (over 900 citations [Web of Science October 24, 2012, Times cited = 1,051]) and widely adopted formalism is still heavily used, ensuring consistency, standardization, and comparability of brachytherapy dose calculations across institutions worldwide.

Since dose calculations based on the TG-43 formalism rely on the superposition of single-source dose distributions over the dwell or seed positions used for treatment, dose distributions can be obtained with minimal calculation time. This fast and practical method facilitates clinical practices such as transrectal ultrasound image-guided live planning and dose distribution optimization. However, the inherent simplicity of the TG-43 formalism can lead to inaccurate dose distributions when the calculation geometry deviates significantly from the reference water sphere used to derive the source’s dosimetric parameters. Generally, TG-43 dose calculation limitations can be attributed to five phenomena: absorption, attenuation, shielding, scattering, and breakdown of the kerma approximation for absorbed dose. Depending on source energy and anatomic site, some or
all of these phenomena may induce significant dose calculation errors. Table 9.1 lists, for both high- and low-energy brachytherapy sources, sites where the formalism leads to significant dose calculation errors (Rivard et al., 2009). The following sections briefly discuss these aspects.

9.3.1.1 Absorbed Dose Differences between Water and Human Tissues

At low photon energies, the mass–energy absorption coefficient $\mu_{\text{en}}/\rho$ varies significantly between tissues due to the importance of the approximately $Z^2$ dependence of the photoelectric cross section. For a given photon energy fluence under conditions of charged particle equilibrium, the absorbed dose-to-water and dose-to-tissue are related by $D_{\text{tissue}}/(\mu_{\text{en}}/\rho)_{\text{tissue}} = D_{\text{water}}/(\mu_{\text{en}}/\rho)_{\text{water}}$. As seen in Figure 9.8, the ratio $(\mu_{\text{en}}/\rho)_{\text{tissue}}/(\mu_{\text{en}}/\rho)_{\text{water}}$ differs from unity for most tissues in the energy range of low-energy sources. At high photon energies, the ratios converge to unity as the Compton cross section becomes more important. This gives rise to significant dependence of the fluence-to-kerma conversion factor on tissue elemental composition and mass density (Landry et al., 2011).

9.3.1.2 Attenuation Differences between Water and Human Tissues

As the linear attenuation coefficient $\mu$ is proportional to $\rho$, density differences between water and human tissues will result in different photon energy fluence and dose distributions, for both low- and high-energy sources. Mass attenuation coefficients $\mu/\rho$ also vary between tissues and water at low photon energies because of the aforementioned dependence of photoelectric cross section on atomic number. As $\mu/\rho$ is larger at low photon energies (disregarding the absorption edges), differences of elemental composition or density will result in larger dose differences for low-energy sources.

### Table 9.1

<table>
<thead>
<tr>
<th>Anatomic Site</th>
<th>Source Energy</th>
<th>Absorption</th>
<th>Attenuation</th>
<th>Shielding</th>
<th>Scattering</th>
<th>Breakdown Dose = Kerma Approximation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Prostate</td>
<td>High</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
</tr>
<tr>
<td></td>
<td>Low</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>N</td>
<td>N</td>
</tr>
<tr>
<td>Breast</td>
<td>High</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>Y</td>
<td>N</td>
</tr>
<tr>
<td></td>
<td>Low</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>N</td>
<td>N</td>
</tr>
<tr>
<td>GYN</td>
<td>High</td>
<td>N</td>
<td>N</td>
<td>Y</td>
<td>N</td>
<td>N</td>
</tr>
<tr>
<td></td>
<td>Low</td>
<td>Y</td>
<td>Y</td>
<td>N</td>
<td>N</td>
<td>N</td>
</tr>
<tr>
<td>Skin</td>
<td>High</td>
<td>N</td>
<td>N</td>
<td>Y</td>
<td>Y</td>
<td>N</td>
</tr>
<tr>
<td></td>
<td>Low</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>N</td>
</tr>
<tr>
<td>Lung</td>
<td>High</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>Y</td>
<td>Y</td>
</tr>
<tr>
<td></td>
<td>Low</td>
<td>Y</td>
<td>Y</td>
<td>N</td>
<td>Y</td>
<td>N</td>
</tr>
<tr>
<td>Penis</td>
<td>High</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>Y</td>
<td>N</td>
</tr>
<tr>
<td></td>
<td>Low</td>
<td>Y</td>
<td>N</td>
<td>Y</td>
<td>Y</td>
<td>N</td>
</tr>
<tr>
<td>Eye</td>
<td>High</td>
<td>N</td>
<td>N</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
</tr>
<tr>
<td></td>
<td>Low</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>N</td>
</tr>
</tbody>
</table>


![Graph](image.png)
applicator materials such as stainless steel can cause deviations. Some applicators also contain high-atomic-number, high-density shielding material such as tungsten, used to protect organs at risk. The most common are shielded vaginal and intrarectal applicators.

9.3.1.4 Scattering Conditions

TG-43 parameters are calculated in water spheres with radii of 15 and 40 cm for low- and high-energy sources, respectively, ensuring that TG-43 calculations represent doses in unbounded water medium over the therapeutically relevant distance range. Situations where full scatter conditions are not met, sources close to the skin, for example, will result in deviation of delivered dose from TG-43. Given the longer pathlength and higher Compton scattering cross section, high-energy photons are more sensitive to geometries with tissue boundaries near the implanted sources. Example clinical geometries include the breast, skin, and lung (Poon and Verhaegen, 2009).

9.3.1.5 Breakdown of the Kerma Approximation for Absorbed Dose

The vast majority of MC studies deriving TG-43 parameters report collision kerma in water, relying on the equivalence of dose and kerma under conditions of charged particle equilibrium. For low-energy sources, electronic equilibrium is reached within 0.1 mm of sources and the kerma = dose approximation is accurate (Ballester et al., 2009). Since electron ranges are larger for high-energy sources, this assumption can lead to dose differences greater than 1% at distances of 7, 3.5, and 2 mm for $^{60}$Co, and $^{137}$Cs, and $^{192}$Ir, respectively (Ballester et al., 2009). Studies performing measurements rarely report data so close to the source. In addition to lack of charged particle equilibrium, beta particles emitted from sources can also cause a breakdown of the kerma = dose approximation. Granero et al. investigated the contributions of both betas emitted from $^{192}$Ir and lack of charged particle equilibrium for distances <2.5 mm from the source’s center (Granero et al., 2011). They found that charged particle equilibrium is reached at 2 mm from the source, and that beta contribution to dose is negligible beyond 2.5 mm. Additionally, CPE breakdown may occur at tissue/high-Z interfaces.

9.3.2 Magnitude of Clinical Impact

Several studies have compared MC dose calculations to the results of the TG-43 formalism. Meigooni et al. were the first to investigate ISA in 1992 using TLD measurements and estimated that dose reductions of 6% could be expected at the edge of a $^{125}$I prostate implant (Meigooni et al., 1992). Chibani et al. and Carrier et al. published results of MC simulations performed in real $^{125}$I and $^{103}$Pd prostate implant geometries in 2005 and 2006, respectively, finding $D_{90}$ reductions of 2–5% due to ISA (Chibani et al., 2005; Carrier et al., 2006). Chibani et al. also investigated the dosimetric impact of the presence of calcifications in the prostate and found $D_{90}$ reductions of up to 37%. Carrier et al. subsequently performed a retrospective study of 28 prostate cancer patients treated with $^{125}$I implants. Tissue effect and Interseed attenuation represent the $D_{90}$ reductions from the TG-43 estimate from each effect. Complete Monte Carlo represents the $D_{90}$ parameter when accounting for both effects. (From Carrier, J. F. 2007. Int J Radiat Oncol Biol Phys 68:1190–1198. With permission.)

Melhus and Rivard showed that tissue effects were negligible for $^{192}$Ir over clinically relevant distances in soft tissue, muscle, and breast using point sources in uniform spherical phantoms (Melhus and Rivard, 2006). This is due to the higher photon energies of $^{192}$Ir. On the other hand, applicator shielding, designed to spare organs at risk, can cause dose reductions of 5–11% for gynecology (GYN) applicators (Markman et al., 2001) and up to 24% for endorectal applicators when compared to the TG-43 predicted dose (Poon et al., 2008). Poon et al. showed that TG-43 dose distributions for $^{192}$Ir breast implants overestimated dose to the skin by 5%, due to differences from full scatter conditions with MC simulations (Poon and Verhaegen, 2009), confirming earlier findings from Pantelis et al. (2005).
9.3.3 Available MC Dose Calculation Tools

9.3.3.1 Monte Carlo Dose Calculation for Prostate Implant

MCPI (Monte Carlo dose calculation for prostate implant) was developed by Chibani and coworkers in 2005 to perform patient-specific prostate implant dose calculations with $^{125}$I and $^{103}$Pd (Chibani and Williamson, 2005). The code is based on the general-purpose MC software GEPTS (Chibani, 1995), using only photon transport for low-energy (<1 MeV) sources. Several variance reduction techniques, in addition to the track-length collision kerma estimator, were developed for MCPI. Photons are transported only once in the source geometry; a phase space file is subsequently used to emit photons from each source surface, while still modeling the source geometry to include ISA. The code handles both voxel (rectangular mesh) and seed (nested cylindrical objects, facilitating analytical ray-tracing) geometries simultaneously using a hybrid geometry model, where voxels intersecting seeds are flagged, thus avoiding the need to query every seed in the geometry when transporting photons, a technique the authors call “voxel indexing.” Finally, MCPI uses ray-tracing instead of analog transport of photons; primary and secondary photon trajectories are projected through the entire mesh of voxels, regardless of whether they have an interaction or not, thereby enhancing the frequency of energy deposition in voxels at larger distances, a technique based upon Williamson’s expected-value track-length estimator (Williamson, 1987). Calculation times are reported to be about 60 s for a $2 \times 2 \times 2$ mm$^3$ voxel mesh (Chibani and Williamson, 2005).

9.3.3.2 PTRAN_CT

The correlated sampling technique (Hedtjarn et al., 2002) has recently been implemented in Williamson’s PTRAN_CT (Dolan et al., 2006) by Sampson and coworkers (Sampson et al., 2012). The PTRAN_CT implementation also included the variance reduction techniques of MCPI described above and expected-value track-length scoring. Dose calculation times of about 3 s for a prostate implant have been reported to achieve 2% statistical uncertainty in a $2 \times 2 \times 2$ mm$^3$ voxel mesh. This code can input digital imaging and communications in medicine (DICOM) CT images and currently uses the EGSnrc CTcreate software to assign cross-section files based upon single-energy CT. An alternative could be the user interface BrachyGUI, developed at McGill University (Poon et al., 2008).

9.3.3.3 BrachyDose

BRACHYDOSE is an EGSnrc (Kawrakow, 2000) user code utilizing Yegin’s multigeometry package (Yegin, 2003) and was developed to perform prostate implant dose calculations (Yegin and Rogers, 2004). The geometry package permits detailed modeling of brachytherapy source geometries, and the code is capable of transporting both photons and electrons from 0.01 to 10 MeV, although the higher energies are not used in brachytherapy. Collision kerma is scored with a pathlength estimator, but dose can also be scored at the cost of longer calculation times. The code has been extensively benchmarked by generating single-source TG-43 parameters for several commercially available $^{125}$I, $^{103}$Pd, and $^{192}$Ir sources (Taylor et al., 2007). BRACHYDOSE is also used for treatment planning of eye plaques with low-energy brachytherapy sources (Thomson et al., 2008). Calculation times of 30 s are reported to produce statistical uncertainties less than 2% for a $2 \times 2 \times 2$ mm$^3$ voxel mesh (Thomson et al., 2010).

9.3.3.4 ALGEBRA

ALGEBRA (algorithm for heterogeneous dosimetry based on Geant4 for brachytherapy) is based on the general-purpose MC code Geant4 and the DICOM radiotherapy (RT) standard. Initially developed by Carrier and coworkers (Carrier et al., 2007), and referred to as Geant4/DICOM-RT in the recent review by Rivard et al. (2010), the code was later modified by Afsharpour and given its current name (Afsharpour et al., 2012). The code can import planning data from a TPS (seed position, air-kerma strength, etc.) and structural data (contours). CT images are imported and a semiautomatic segmentation method uses the CT calibration curve to assign densities and organ contours to assign elemental compositions to each voxel. Seed geometries and voxels do not overlap in ALGEBRA; voxels intersecting seeds are discarded, and the missing volume is replaced by water. The new layered mass geometry functionality of Geant4, where the overlap of seeds and voxels is permitted, should eliminate the need to perform this step (Enger et al., 2012a). Dose scoring is performed using the “parallel world” functionality of Geant4 by creating a scoring mesh independent of the transport geometry. The resolution of the CT-derived transport geometry can be modified, as well as that of the scoring mesh, and they need not be equivalent. Secondary electrons are immediately discarded and collision kerma is scored with a pathlength estimator. Both kerma in water or medium by transporting photons in medium, $K_{w,m}$ and $K_{m,m}$ can be scored. The use of source phase space files is also possible. Calculation times in ALGEBRA for $2 \times 2 \times 2$ mm$^3$ voxels are about 6 and 12 min for 2% statistical uncertainty for a breast $^{103}$Pd implant and a prostate $^{125}$I implant, respectively.

9.3.4 Issues Involved in Clinical Implementation

9.3.4.1 $D_{w,m}$ or $D_{m,m}$

An excellent overview of this issue can be found in the report of AAPM Task Group 186 (Beaulieu et al., 2012). MC simulations provide the ability to calculate the energy deposition in the medium present in each voxel of a treatment geometry, either from photons ($K_{m,m}$) or from electrons set in motion by photons ($D_{m,m}$). An alternative approach to reporting dose to medium would be to calculate the energy deposition in a unique reference medium. For historical/practical reasons, this reference medium has consistently been water. While in external beam radiotherapy (EBRT) $D_{w,m}/D_{m,m}$ differences between soft tissues and water are of the order of 2%, at the energies of brachytherapy sources these differences can be much larger. In EBRT, conversion is performed...
by using ratios of mass collision stopping power under the assumptions of Bragg–Gray cavity theory (electron ranges larger than size of cavity). For brachytherapy, the method of conversion depends on the cavity size. Millimeter-sized voxels are several times larger than the range of secondary electrons produced by photons emitted from low- or high-energy sources; thus, ratios of mass energy absorption coefficients \( \mu_{\text{en}}/\rho \) (large cavity theory) may be used to convert \( K_{\text{n,m}} \) to \( K_{\text{en,m}} \). The \( \mu_{\text{en}}/\rho \) differences between tissue and water are generally important for low-energy sources (up to 80% for certain soft tissues), while the situation is similar to EBRT for high-energy brachytherapy sources (3–5%).

Complications arise when we consider that the radiation response of tissues and tumors correlates with the energy deposited in the nuclei of cells. The dimensions of mammalian cells (~10 \( \mu \text{m} \)) mean that they act as Bragg–Gray or small cavities at the energies of high-energy brachytherapy sources and as intermediate cavities bracketed by large and small cavity theories at the energies of low-energy brachytherapy sources. This makes the conversion of \( D_{\text{n,m}} \) to \( D_{\text{en,m}} \) using \( \mu_{\text{en}}/\rho \) questionable for low-energy sources and means that different conversion methods are required across the brachytherapy energy range. There is also no clear indication that water is the best material to represent the nuclei of various cells (Enger et al., 2012b). For these reasons, it has been recommended to report both \( D_{\text{n,m}} \) and \( D_{\text{en,m}} \) when performing MC dose calculations for brachytherapy (Beaulieu et al., 2012).

### 9.3.4.2 Tissue Composition Selection

MC dose calculations generally require voxel-by-voxel assignment of tissue density and elemental composition (mass fraction of each element composing the tissue). While tissue composition essentially plays no role for brachytherapy MC dose calculations with high-energy sources, it becomes very important for low-energy sources due to the \( Z \) dependence of \( \mu/\rho \) and \( \mu_{\text{en}}/\rho \) (photoelectric effect). While tissue density also affects dose distributions, it is readily obtained from CT images of the treated anatomy using a Hounsfield unit (HU) to density calibration (similar to the HU to electron density calibration used in external beam).

Soft tissue composition, on the other hand, is difficult to correlate with HU, as several human tissues possessing different compositions have HU falling in the \([–100, 100]\) range around water (Schneider et al., 2000). In most MC brachytherapy studies, tissue composition is assigned on the basis of anatomical organ contours, that is, prostate tissue is uniformly assigned to the prostate contour and breast tissue is uniformly assigned to the breast contour. With this practice, the question of what constitutes prostate or breast tissue arises. Literature reports (ICRU, 1992; ICRP, 1975) provide compositions for a wide range of human tissues; however, these compositions are based on old (1930–1970) measurements on a small number of subjects and may be of questionable accuracy in characterizing average tissue composition in the population served by brachytherapy. Some reports provide a range of composition for a given tissue, an indication of intersample variability (Woodard and White, 1986). Landry et al. investigated the sensitivity of dose distributions from low-energy sources for tissues of interest in prostate and breast brachytherapy using clinical \(^{125}\text{I} \) and \(^{103}\text{Pd} \) implants, respectively, and postimplant CT images (Landry et al., 2010). For prostate, two compositions were found in the literature resulting in \( K_{\text{n,m}} D_{\text{en}} \) variations of 3.5%. For breast, which is composed of adipose and glandular tissue, Landry et al. generated a uniform tissue. The variation of the proportion of each tissue from 70/30 to 30/70 yielded \( K_{\text{n,m}} D_{\text{en}} \) variations of 10% (or about 6% when considering \( K_{\text{en,m}} \)). For a given adipose/gland proportion, the variability of each tissue’s composition resulted in dose variations of 10%. Recent studies on breast brachytherapy with low-energy sources suggest segmenting adipose and mammary gland on CT images, as the uniform tissue approximation may fail to reproduce the dose distributions obtained from segmented geometries (Sutherland et al., 2011; Ashfarour et al., 2012). Only one modern study of elemental composition of tumors appears to exist, that of Maughan et al. (1997), which demonstrated a large variation in carbon content (8–32% by weight) and mineral ash (0.9–3% by weight), which gives to variation in \( (\mu_{\text{en}}/\rho)_{\text{tumor}} \) of 20% at 30 keV.

### 9.3.5 Image Usage

#### 9.3.5.1 CT

CT images are generally the standard input to MC calculation, as CT image intensity approximately tracks the relative linear attenuation coefficient, evaluated at the effective scanning energy, of the underlying tissue. Converting HU to electron or mass density is a relatively straightforward step, as explained above. For low-energy sources, the issues of the previous section demand special attention, but for high-energy sources, where tissue composition effects are essentially negligible, CT images provide all the necessary information to accurately calculate the dose. Additionally, the location of implanted seeds, catheters, or applicators is easily derived from CT images.

CT artifacts, especially those caused by the high-density components of implanted LDR seeds (streaking artifacts) and other foreign metal bodies implanted in the body, can seriously degrade the information required for accurate dose calculations. Metal artifact correction algorithms exist and can alleviate this problem (Xu et al., 2011). However, metal artifact reduction cannot be considered to be a solved problem, especially at the level of quantitative accuracy needed to specify tissue composition for low-energy brachytherapy. Even relatively subtle streaks can degrade voxel-by-voxel single-energy CT tissue identity analyses (Schneider et al., 1996).

#### 9.3.5.2 MRI

With its excellent soft tissue contrast, magnetic resonance imaging (MRI) permits the identification of lesions and organs at risk not visible on CT, and currently plays an important role in GYN brachytherapy (Potter et al., 2008). By differentiating tissues, it permits a better distribution of radiation dose at the time of planning. Prostate delineation is also acknowledged as being superior on magnetic resonance (MR) compared to CT imaging (Viswanathan et al., 2007). However, the lack of density information currently limits the use of MR images for MC dose calculations. Work has
been performed to employ MR images in EBRT treatment planning and for PET attenuation correction, and two types of approaches have been developed (Johansson et al., 2011). Methods based on anatomy attempt to deform patient MR images to a reference MR image, and apply the resulting deformation on a corresponding reference CT image (Kops and Herzig, 2007). Voxel-based approaches attempt to classify voxels of MR images into tissue types (Keereman et al., 2010). Following classification, bulk electron density assignment or direct conversion to HU or electron density is performed. These techniques have not been evaluated in the context of brachytherapy dose calculations so far. For prostate EBRT, Lee et al. reported dose calculation errors of less than 2% when segmenting images into bone and soft tissue compared to using the full CT (Lee et al., 2003). Given that high-energy photons employed in brachytherapy are relatively insensitive to tissue type assignment (see Table 9.1), we can expect similar results. For low-energy photons, the added complexity of assigning a correct tissue composition would need to be considered. Geometric distortion of MR images due to magnetic field inhomogeneities, gradient non-linearity, susceptibility effects, and chemical shifts have also been considered in EBRT (Lee et al., 2003). Brachytherapy is possibly less sensitive to distortions than EBRT, as the distortions are generally acknowledged to be low in the center of the field of view (FOV), where MR compatible applicators or seeds would be positioned.

9.3.5.3 Ultrasound

Transrectal ultrasound images are used to guide needle implantation for LDR seed implants and catheter implantation for HDR brachytherapy of the prostate. Ultrasound images provide information on the location and dimensions of certain organs of interest and permit TG-43 dose calculations (Chew et al., 2009). While ultrasound-based tissue typing is an active area of research (Zhou et al., 2009), it does not yet provide patient-specific density and tissue composition information necessary to perform MC simulations. Additionally, the presence of the TRUS probe deforms the organs of interest. External transducers, permitting non-invasive image acquisition, can avoid this issue but are currently not used in brachytherapy applications. One approach could be to assign generic densities and compositions to organ contours, as explained above for MR images. Using the TRUS-derived seed positions, ISA could also be modeled in prostate implants.

9.4 Future Use of Monte Carlo Methods for Brachytherapy

Like the previous two sections, opportunities for future applications of MC methods for brachytherapy may be divided into two categories: (a) single-source dosimetry for preclinical or research purposes and (b) patient-oriented applications such as for treatment planning.

9.4.1 Single-Source Dosimetry

Over the past decade, there have been significant advances in characterizing radiation dose distributions for individual brachytherapy sources. These studies have examined the sensitivity of phantom composition on absorbed dose, the influence of motion of dynamic internal components within the source, effects of manufacturing variations on resultant dose distributions, and previously ignored radiations such as beta penetration and characteristic x-ray influence. Forthcoming advances upon using MC methods for brachytherapy may include assessment of dosimetric contributions from radioimpurities, characterization of dissolvable brachytherapy sources, design of multidrug sources, integration of brachytherapy dosimetry and patient imaging, and of course evaluation of new radionuclides. Since the focus can range from methodical improvements on existing paradigms to highly creative blue sky ideas, the possibilities seem endless.

9.4.2 Treatment Planning

The first decade of the twenty-first century has seen the exciting birth of image-based MC treatment planning for brachytherapy. Early advances have examined MC code variance reduction techniques while maintaining calculation accuracy, patient tissue assignment and sensitivity of results to such assignments, consideration for dose specification choice such as dose to medium in medium or dose to water in medium, and retrospective studies trying to correlate clinical outcomes with the more accurate dose distributions using MC methods.

Future directions for MC-based brachytherapy treatment planning will continue the organized improvements on existing methods, integration of MC computations during the dose optimization phase, use of MC techniques in inverse optimization, resolution of imaging artifacts such as from seeds or high-Z applicator, correcting for limited image datasets as needed to provide full radiation scatter conditions to simulate the clinical environment, and accounting for changes in brachytherapy applicator positioning relative to the relevant tissues in the patient. Perhaps MC methods can even play a role in accounting for organ motion during treatment or between treatments, as has been briefly explored for external beam irradiation. Multicenter brachytherapy dosimetry comparison trials versus MC as the gold standard are now just on the horizon. Radiobiological modeling on the submicron scale for microdosimetry and improved understanding of RBE effects, as well as source design for optimal exploitation of RBE effects, are also of interest.

Given the increased interest in brachytherapy dosimetry and the improved coordination of international research teams, the future is bright for using MC methods in the field of brachytherapy.

References


10.1 Introduction

Advanced radiation therapy (RT), such as intensity-modulated radiation therapy (IMRT) (Purdy, 1996; Burman et al., 1997; Webb, 1998; Boyer and Yu, 1999) and intensity-modulated rotational therapy, including volumetric modulated arc therapy (VMAT) and RapidArc (Lagerwaard et al., 2008), has been widely accepted as the standard treatment strategy for many treatment sites in the field of radiation oncology because of the ability to provide quality conformal dose distributions. Advanced RT treatment techniques offer better sparing for the surrounding normal tissues than conventional treatment methods and, therefore, lead to less normal tissue complications (Salama et al., 2006; Studer et al., 2008; Zelfskey et al., 2008) and the possibility for dose escalation to the treatment target (Al-Mamgani et al., 2009). To provide a conformal dose distribution, advanced RT treatment techniques require a more complex treatment planning process utilizing computers in the plan optimization and a more complex beam delivery system, which may utilize a multileaf collimator (MLC) with complex leaf motion sequences. For intensity modulation purposes, the radiation field is created with a number of small beamlets with a typical size of $10 \times 10$ or $5 \times 5\, \text{mm}^2$ delivered through many small, irregular, and asymmetric MLC fields, which obscure the relationship between the accelerator monitor unit (MU) setting and the radiation dose received by the patient. When on-line images or other localization and tracking systems are used for treatment guidance and intervention, the beam delivery may become even more complex. Furthermore, patient anatomy heterogeneity, organ motion, and deformation may add additional uncertainties to the actual dose distribution received by the patient. Overall, potential errors associated with advanced RT include dose calculation inaccuracies, plan transfer errors, beam delivery errors, and target localization uncertainties due to patient setup errors and organ motion during the treatment. Considering the serious consequences of these errors, a comprehensive quality assurance (QA) should be performed before and/or during the patient treatment.

The ultimate goal of RT QA is to ensure that the patient will receive the planned dose distribution. However, in clinical routine practice, only a part of the radiotherapy process is usually verified. Techniques, methods, and protocols for RT QA vary from facility to facility in current practices. These methods can be categorized as measurement based, calculation based, and simulation based. The measurement-based method includes measuring and verifying the fluence map for all fields using a film or a two-dimensional (2D) detector array (Grein et al., 2002; Greer and Popescu, 2003; Iori et al., 2007; Pallotta et al., 2007) to measure and compare point doses or a 2D dose distribution using an ion chamber or a 2D detector array in a dummy phantom (Dong et al., 2003; Ma et al., 2003). The measurement-based method can verify the treatment plan, dose calculation in the phantom, and plan transfer and delivery. However, it is unable to verify the treatment dose to the patient, especially for a treatment site with heterogeneity in which the absolute dose distribution can be significantly different from that in a homogeneous phantom. In addition, patient setup error, organ motion and deformation, and their effects on the patient’s dose distribution cannot be adequately evaluated since no real-time imaging is available to record the variation of the patient geometry during treatment. Other concerns include the requirement of treatment machine time and the off-hour workload to perform
these QA measurements. The calculation-based method can be performed in the patient geometry as specified by the patient’s computed tomography (CT) scan or other image data set. This technique can be performed at any time using computers, making it less labor intensive. Most calculation-based QA methods currently used clinically employ simple dose/MU calculation algorithms. The effects of patient heterogeneity, setup error, and organ motion on the dose distribution are not included in the calculation. They are frequently used as a secondary MU check tool. In addition, plan transfer and beam delivery cannot be verified by this method. Computerized Monte Carlo simulation can be used as a comprehensive method for RT treatment QA because it can simulate the measurement by performing numerical experiments in the patient geometry and the treatment machine head based on fundamental physics principles. The simulation-based method can perform many tasks that cannot be done with measurement-based methods, such as determining the dose in a patient, and it can provide more information in an accurate manner. Together with measured or recorded beam delivery information, it can verify the plan transfer and treatment delivery as well.

The Monte Carlo method is a well-known computer simulation method utilizing random sampling techniques for particle transport and interaction with material. Its accuracy for radiation dose calculation in a complex and heterogeneous geometry has been extensively benchmarked (Rogers and Bielajew, 1990; Rogers, 2006; Andreo, 1991). Compared with those correction-based and even model-based methods used in conventional algorithms, the Monte Carlo method can model the details of the beam delivery system more precisely because each individual particle is simulated separately in the accelerator head and patient geometry (Rogers et al., 1995). The computation speed of the Monte Carlo method is always a major concern. However, the recent developments in computer techniques and efficiency-enhancing techniques have made it practical for clinical use. Therefore, the Monte Carlo method can be employed as an accurate and comprehensive QA tool for advanced IR applications.

The following sections will discuss Monte Carlo dose calculation, techniques and implementations, and the clinical applications of Monte Carlo-based QA for advanced radiotherapy treatments.

10.2 Techniques and Implementation Methods

10.2.1 Overview of Monte Carlo-Based QA
Monte Carlo dose calculation is performed by transporting the primary radiation particles and their descendants through the patient phantom geometry and recording their energy deposition along the path with a random sampling-based computer simulation process. One can perform Monte Carlo-based patient dose calculation for IR starting with the electrons exiting from the accelerator vacuum window and following them and their descendants through the rest of the accelerator head including the beam collimation and monitoring devices and then the patient geometry. This method is very inefficient, since most of the particles are blocked by the collimators. Realistically, the radiation beam can be represented by a phase space data set (Rogers et al., 1995) or a source model (Ma et al., 1997) at the plane below the fixed components of the accelerator head. The phase space data is generated by direct Monte Carlo simulation of the patient-independent components in the accelerator head and the source model parameters are derived from the phase space data (Ma et al., 1997; Ma and Jiang, 1999; Chetty et al., 2000; Deng et al., 2000; Fix et al., 2001c; Chetty, 2007) or measured beam data (Jiang et al., 2001). The phase space data and the source model can be used for all patients to be treated on this accelerator. Properties of the particles, including type, energy, direction, and location can be retrieved from the phase space data file or reconstructed from the source model. Monte Carlo-based RT QA can start from the phase space plane or the source model and follow all the particles and their descendants through the patient-dependent accelerator components (e.g., jaws, compensators, and MLC) and the patient geometry. This is referred to as direct Monte Carlo simulation of the beam delivery. Because many particles will be blocked by the beam collimation system, direct Monte Carlo simulation for advanced radiotherapy QA is not efficient. An alternative method is to use the beam intensity map. Instead of simulating a particle through the collimation system, the particle is assigned a weighting factor based on the probability for it to go through the collimation system. The distribution of the weighting factor is defined by the transmission properties of the compensator or the MLC leaf geometry and the sequence of the leaf movements. The particle’s weighting factor will determine the particle’s contribution to the dose. The intensity map can be obtained from the 2D fluence measurement or reconstruction from the treatment delivery information as discussed in the following sections.

10.2.2 Treatment Delivery Information
As illustrated in Figure 10.1, in addition to the phase space file/source model and the patient geometry as specified by a CT or other imaging data set, the treatment delivery information is required by the Monte Carlo-based QA process. First, the radiation beam and its direction relative to the patient geometry need to be established. The phase space data or source model will be selected based on the actual radiation beam used for the treatment. The beam direction is usually described using the gantry, collimator, and couch angles for a conventional medical linear accelerator. It is sometimes described differently for other treatment machines, such as the nodes for the CyberKnife or the rotation angles for a tomotherapy machine. This information can be found in the treatment plan, record and verify (R&V) system, and the treatment log file, as illustrated in Figure 10.1. The jaw setting or MLC leaf positions, which define the shape of a radiation field or an MLC field segment, and the MUs used for the radiation field are required by the direct Monte Carlo simulation and the intensity map reconstruction program for the entire treatment. The MU information can be found in the
Monte Carlo as a QA Tool for Advanced Radiation Therapy

10.2.3 Direct Monte Carlo Simulation

Direct Monte Carlo simulation means to simulate the particle transport and interaction in the patient-dependent field-defining components in the treatment delivery system together with patient dose calculation using the Monte Carlo method. The radiation particles and their descendants are transported in the beam collimation system and the patient geometry until they exhaust their energy or escape to the free space. Geometry parameters of the physical components of the treatment delivery system are determined based on the treatment delivery sequence (jaw setting, MLC leaf positions, and MUs for individual segments) and the geometry of machine components. During the simulation, the beam collimation geometry can be considered to be stationary at any particular moment (treatment segment). This works for “stationary” beam delivery, for example, conventional radiotherapy and step-and-shoot IMRT, and for “dynamic” beam delivery, such as dynamic IMRT and intensity-modulated arc therapy (IMAT)/VMAT/RapidArc. The particles are weighted by the MUs of each treatment segment and, thus, produce “absolute” rather than “relative” dose distributions. Various beam modifiers, such as the block, wedge, and MLC, have been incorporated in the patient simulation geometry for IMRT dose verification studies by several research groups (Li et al., 2000; Fix et al., 2001a,b; Keall et al., 2001; Kim et al., 2001; Liu et al., 2001; Spezi et al., 2001; Verhaegen and Liu, 2001; Aaronson et al., 2002; Siebers et al., 2002; Heath and Seuntjens, 2003; Leal et al., 2003; Fragoso et al., 2009). With direct Monte Carlo simulation, the details of the beam collimation device, such as the tongue-and-groove structure and the leaf end shape of the MLC, can be considered in the patient dose calculation accurately if the treatment delivery system is modeled precisely. However, direct Monte Carlo simulation of the particle transport through the collimator system is time consuming because most particles are blocked before they can reach the patient geometry. To improve the simulation efficiency, a variety of strategies have been applied. One method is to terminate the particle transport in the beam collimation system using high-energy cutoffs.
for photon and electron particles. Another method is to disre-
gard the secondary Compton scattering photons (Siebers et al.,
2002) or the secondary electrons (Tyagi et al., 2007) because they
deposit most of their energy locally. Some researchers have also
investigated methods to only model the radiation transport at
the field edge to account for the effects of partial transmission
Significant time saving has been observed using these methods.

10.2.4 Beam Intensity Map for Monte
Carlo-Based QA

As mentioned above, instead of performing direct Monte Carlo
simulations, the beam intensity maps obtained by direct mea-
surements or analytical reconstruction methods can be used in
Monte Carlo patient dose calculation.

10.2.4.1 Beam Intensity Map Measurement

The beam intensity map for each beam angle can be mea-
sured using a film, EPID, or 2D detector arrays, such as Map
CHECK (Sun Nuclear, Melbourne, FL) and MatriXX (IBA,
Schwarzenbruck, Germany). Measurements can be performed
at various locations, for example, the beam exiting window, the
isocenter plane or any extended distances, pretreatment without
a patient in the beam, or during the treatment with a patient in
the beam. All the measured intensity maps without a patient in
the beam can be used for dose calculation after some corrections
are made (e.g., the detector response and back scattering from
surrounding materials). The map measured on-line behind the
patient is called the transmission (transit) intensity map, which
also includes the effect of beam attenuation and scattering by
the patient. This effect must be corrected or removed from the
transit intensity maps before they can be used for dose calcula-
tion. Ray-tracing algorithms are frequently used to derive the
entrance intensity map in front of the patient from the transit
images. Overall, a careful calibration of the 2D detector array
and proper consideration of the patient scattering and attenu-
ation will improve the accuracy of the patient dose distribution
calculated this way. Some investigators have reported the use of
the EPID for routine machine QA, pretreatment dose verifica-
tion, and in vivo dosimetry (Boellaard et al., 1998; Kroonwijk
et al., 1998; Pasma et al., 1999a,b; van Zijtveld et al., 2006; van
Elmpt et al., 2008).

10.2.4.2 Intensity Map Reconstruction

In addition to the use of direct measurements, the beam intensity
map can also be reconstructed based on the treatment delivery
information. Reconstructing a 2D intensity map based on the
recorded MLC leaf sequence and MU information for advanced
radiotherapy dose calculation can save a great deal of simulation
time compared with direct Monte Carlo simulations and can
be more accurate and versatile compared with on-line intensity
map measurements because the corrections are patient specific.
The beam delivery information, including MU and MLC leaf
sequences, can be derived from various sources as described in
Section 10.2.2.

The 2D beam intensity map can be built at the midplane of
the MLC and can be expressed at the isocenter plane for con-
venience extending over the maximum treatment field dimen-
sions. The pixel size of the beam intensity map may affect the
simulation results of the patient dose distribution. It was shown
(Li et al., 2010) that the pixel size should be as small as 0.2 mm.
Figure 10.2 demonstrates that the MUs for each MLC segment
are added to the beam intensity values of all pixels in the open
area of an MLC segment (particle B, D, and H). Variable frac-
tions of the MUs are added to the beam intensity values of all
affected pixels, based on the attenuation and scattering effect of
the MLC leaves, for example, using the effective transmission

![Diagram showing how the intensity map is built based on the leaf sequence information to include the effects of photon partial
transmission through the MLC leaf end and the tongue-and-groove geometry. A through J are photons emitted from the point source and the
extra-focal source.](image-url)
factors of the MLC leaf (particle A, C, E, and J) and the MLC leaf tongue-and-groove structure (particle F, G, and I). The dimensions of the leaf opening are set according to the treatment beam information plus an offset to consider the partial transmission effect of the leaf end or leaf side. The leaf offset can be calculated based on a picket-fence film measurement as described in the literature (Li et al., 2010). The transmission factor for the areas under the photon jaws can be assumed to be zero. The dimensions of the photon jaw opening are set according to the actual treatment setups for each MLC segment. The leaf sequences are read sequentially and the beam intensity map is built for every beam direction. It was shown (Li et al., 2010) that it is essential to carefully consider the effects of the details of the beam delivery system, including the source distribution, MLC thickness, MLC transmission and scattering, MLC tongue-and-groove structure, and the leaf end shape. These factors can affect the doses at various field locations differently due to the complexity of the MLC leaf sequence and, thus, they can result in not only different mean doses but also different dose uncertainties at individual field locations. Their effects are case dependent because the MLC leaf sequence is specific to the treatment plan of a particular patient.

10.2.5 Monte Carlo Dose Calculation

During a Monte Carlo simulation, the spatial source distribution and its effects can be considered by sampling particles from multiple subsources, which include a primary point source at the target location, an extra-focal source to account for photon scattering in the primary collimator and flattening filter (Jiang et al., 2001) and an electron contamination source at a location above the jaws to account for the electron component in the beam (Yang et al., 2004). The MLC collimation can be considered either by a direct Monte Carlo simulation of particle transport and interaction in the MLC leaf geometry or by ray tracing through the MLC leaf geometry to determine whether the particles can go through the MLC opening (Li et al., 2010).

When the beam intensity map was initially implemented for Monte Carlo-based dose calculation, the weight of a phase space particle was altered based on the value of the pixel in the intensity map through which the particle was traveling (Ma et al., 1999, 2000). The gantry/couch/collimator angles were changed automatically after the simulation of each treatment field. A 3D voxelized rectilinear phantom converted from the patient/phantom CT scan (Ma et al., 1999; Chetty et al., 2007) has been used for the dose calculation. The density and material for each voxel are converted based on the CT number. The conversion function for any individual CT scanner is unique and is obtained based on the CT calibration with a standard phantom. To avoid the spatial averaging effects (Ai-Dong et al., 2005), a proper voxel size needs to be used when heterogeneity presents. Since the Monte Carlo method is based on random sampling, the calculated doses always contain statistical uncertainties. The doses calculated using analytical methods do not have statistical uncertainties; however, they usually contain larger systematic errors in heterogeneous phantoms compared to the Monte Carlo method. By simulating a sufficient number of particles, the statistical uncertainty of a Monte Carlo result can be reduced to such a degree (e.g., <2%) that will be considered clinically insignificant. Such a result will be useful and meaningful clinically. Otherwise, large statistical uncertainties may lead to confusing results as shown by some researchers (Jeraj and Keall, 2000; Jiang et al., 2000).

Another issue that is often encountered in the patient dose calculation is whether the dose should be reported as dose-to-water or dose to the local medium. By default, the initial calculated Monte Carlo dose is dose to the local medium. Compared with the results calculated by conventional methods, significant differences are usually seen since conventional methods are reporting dose-to-water by default. If one converts the dose value calculated by Monte Carlo simulation of millivolt photon beams in a bone voxel to dose-to-water using the mass electron stopping power ratio between bone and water, more significant differences may result in the dose value (Siebers et al., 2000, Ma and Li, 2011). It is still debatable as to whether to use dose-to-water or dose-to-medium for advanced IR using megavoltage photon and electron beams since it depends on the actual situation (e.g., the particle type and treatment depth).

10.3 Clinical Applications

The radiation treatment procedure for a patient includes treatment planning using a treatment planning system (TPS), treatment plan transfer, and verification using an R&V system, and treatment delivery on a clinical treatment machine. Accordingly, as shown in Figure 10.3, the Monte Carlo method can be applied in the TPS commissioning and verification, patient-specific plan QA, and on-line/off-line patient treatment QA.

10.3.1 MC for TPS QA

Since computers were employed in IR treatment planning, many dose calculation algorithms have been developed, which become more accurate but also more sophisticated and computation intensive. Those early correction-based algorithms can compute the patient dose distributions based on the dose data measured in water for a clinical treatment machine with simple corrections for the differences in beam attenuation and scattering between a uniform water phantom irradiated by radiation beams with regular field sizes and the heterogeneous patient geometry irradiated by the actual treatment beams with irregular field shapes. More recent model-based algorithms predict patient dose distributions from primary particle fluence and a dose kernel, which can be derived from direct measurements, analytical calculations, or Monte Carlo simulations. The accuracy of these algorithms as implemented in different TPS depends critically on the details of the beam fluence modeling for individual treatment delivery systems and the kernel variation to account for the effect of particle attenuation and scattering in the patient’s heterogeneous geometry. The same applies to the Monte Carlo
dose calculation method, in which the dose calculation accuracy can be affected by the implementation details such as the particle phase space reconstruction and patient geometry setup. To ensure the accuracy of the dose calculation algorithms of a TPS, a series of tests must be performed during the commissioning process before its clinical use, which has been summarized by Anders Ahnesjö and Maria Mania Aspradakis in their review paper on external photon beam dose calculation (Ahnesjo and Aspradakis, 1999).

Source modeling and beam commissioning are important elements of TPS commissioning, especially for model-based algorithms and Monte Carlo techniques. The source can be modeled as a single point source or multiple subsources in which the source size and spatial source distribution can be considered in the dose calculation. One application of Monte Carlo-based QA is to evaluate the accuracy of the source modeling and commissioning by comparing the dose distributions calculated by the TPS and the Monte Carlo code for some representative fields defined by the jaws and/or MLC. Percentage depth dose (PDD) distributions for different square fields, lateral dose profiles at different depths, and beam output factors for different field sizes are usually used for the evaluation.

The dose calculation algorithms implemented in a TPS can be verified by standard phantom measurements. The QA measurements for a typical patient’s treatment plan are usually performed at a reference point with an ionization chamber and a plane with a film or MLC. Percentage depth dose (PDD) distributions for different square fields, lateral dose profiles at different depths, and beam output factors for different field sizes are usually used for the evaluation.

Although some 3D detector arrays and image-based measurement techniques (e.g., gel dosimetry) have been developed. It is also difficult to mimic the real patient geometry using a heterogeneous phantom. In vivo dosimetry is typically performed with detectors on the patient’s surface and cavities or inserted invasively in the patient’s body. Corrections have to be made to the detector readings to account for the variation of detector response and other factors due to noncalibration conditions. In comparison, the Monte Carlo method can be used to simulate the 3D dose distribution in the patient geometry represented by the patient’s CT or in a heterogeneous phantom that can represent the patient geometry accurately. The dose volume histograms (DVH) for the treatment targets and organs at risk (OAR) can be compared together with the isodose distributions between the TPS and the Monte Carlo method. This comparison can also be used to verify the MUs for the patient’s treatment plan as calculated by the TPS, which serves as a part of the independent physics check in the treatment planning QA process. Several research groups have published their results of this application for different TPS (Ma et al., 2000; Li et al., 2001; Weber and Nilsson, 2002; Francescon et al., 2003; Wang et al., 2006; Bush et al., 2008; Sarkar et al., 2008).

The details of the beam delivery system and their effects on the patient dose distributions may not be precisely accounted for in the TPS dose calculation due to simplification in the implementation and uncertainties resulting from the commissioning procedure. The TPS usually employs a global correction factor (a fudge factor) based on the mean ratio of measured point doses for a number of test plans to those predicted by the TPS calculation. This correction factor can only reduce but not eliminate the mean dose difference because this point dose ratio is case dependent and position dependent as a result of the leaf sequence for a specific treatment plan. The Monte Carlo method is more advantageous in the verification of the absolute dose and
for determination of the correction factor since the mean dose ratio can be evaluated based on the dose in a region rather than at one or more points.

### 10.3.2 MC for Patient-Specific Plan QA

Dose measurement in a phantom is frequently used to verify the dosimetry accuracy of the phantom plan, plan transfer, and machine delivery for a specific patient. However, this plan QA process only ensures the dosimetry accuracy of the phantom plan, which is calculated based on the patient’s treatment plan and the phantom geometry. This process does not really verify the dose to be received by the patient. Monte Carlo dose calculation based on the patient geometry and the treatment delivery information can verify the dose distribution to be received by the patient and, thus, it can be utilized as a patient-specific QA tool. The treatment delivery information obtained at the various steps of the radiotherapy process can be used for the Monte Carlo QA calculation. A good agreement between the patient dose distribution predicted by the TPS and that by Monte Carlo simulation has been observed for various treatment sites and different treatment planning systems. However, significant differences in the doses to the target and the critical structures have been observed for some treatment cases, especially in the region where heterogeneity exists. The dose difference is mainly caused by the systematic error of the simple dose calculation algorithm in the heterogeneous region and inadequate consideration of the effects of the beam delivery system details such as the leaf thickness, the MLC leaf end, and tongue-and-groove structure. The faults in the implementation and programming of the treatment optimization, dose calculation, and leaf sequencing algorithms can, of course, produce a significant dose error. Many researchers have proposed and demonstrated their methods for Monte Carlo-based patient dose verification. For example, Ma et al. (2000), Pawlicki and Ma (2001), Wang et al. (2002), Li et al. (2004), Yang et al. (2005), and Jiang et al. (2006) reconstructed the intensity map using the leaf sequence file or radiotherapy treatment plan (TPP) file from the TPS. Luo et al. (2006) used the treatment log file and R&V record and Lin (Lin et al., 2009) used the measured EPID images for intensity map reconstruction. Their results show that the Monte Carlo method is a useful and practical tool for patient dose verification. Together with measurement or recorded beam delivery information, the treatment delivery and plan transfer can also be verified.

On the basis of the patient’s four-dimensional (4D) CT data, dose calculation can be performed for each phase of the breathing cycle. After performing image deformation and registration for all the breathing phases, the dose can be summed up for the entire treatment with a proper weighting factor based on the patient’s average breathing cycle and, thus, a 4D dose distribution can be obtained (Keall et al., 2004). The time association between the beam-on time and the breathing phase can be associated with the real-time positioning management (RPM) system or other respiratory tracking signals. Although dosimetric uncertainties still exist due to the uncertainties of the 4D CT images, the correlation between the external surrogate and internal organ/target motion, and the deformation registration of CT images at various phases, 4D dose calculation using the Monte Carlo method is still valuable for the investigation of the dosimetric effects of the patient’s intrafractional motion and organ deformation.

### 10.3.3 MC for On-Line and Off-Line Treatment QA

Utilizing a film, 2D detector arrays, EPID, and 3D patient imaging for pretreatment or on-line treatment measurement, the Monte Carlo method can be used for pretreatment, and on-line and off-line patient treatment QA for advanced RT. As described in the previous sections, a film, 2D detector arrays, and EPID can be used to measure the entrance fluence map in the absence of the patient or to reconstruct the entrance fluence map based on the transit intensity map measured with the patient in place during the treatment. The entrance fluence map as directly measured by a film, 2D detector arrays, and EPID must be corrected for the detector response variation and other effects due to attenuation and scattering by the measurement device itself. The entrance fluence map can be derived from the transit intensity maps by ray tracing after correcting the beam attenuation and scattering from the patient, or reconstructed based on the MLC leaf positions and MUs recorded by the R&V system or the treatment log file. Subsequently, on-line or off-line dose calculation using the Monte Carlo method can be performed based on this information. Patient CT images acquired with the in-room CT (cone-beam or CT-on-rails) or other 3D imaging systems (ultrasound and MRI) before and during the treatment can be used for the on-line dose calculation and the calculated dose distribution can be used to guide the treatment. Adaptive radiotherapy is possible if the dose calculation and image processing is sufficiently fast. The advances of computer techniques and variance reduction techniques have accelerated the Monte Carlo dose calculation and image processing dramatically in the recent years. The overall computation time for this procedure can be within several minutes. One example is the COMPASS system developed by Boggula et al. (2011), which reconstructed the 3D dose distribution for treatment verification using the fluence maps measured off-line or on-line with a 2D detector array system. Other systems similar to this can be employed for off-line or on-line patient dose verification and adaptive radiotherapy.

### 10.4 Conclusions

Advanced RT techniques have received widespread clinical applications for cancer treatments. A comprehensive and practical QA program is essential to ensure the dosimetry accuracy of treatment planning, plan transfer, and treatment delivery for advanced RT. The Monte Carlo method can serve as a useful QA tool in the various steps of the advanced RT process. Techniques and methods to perform Monte Carlo-based QA have been discussed in this chapter. With these methods, a patient’s treatment can be verified specifically with confidence. Off-line and on-line
dose verification can be performed with on-line imaging and real-time measurement techniques, which, in turn, can provide useful information for the oncologist to make on-line or off-line treatment adjustments. The Monte Carlo calculated dose distributions incorporated into the patient’s setup and organ motion/deformation information during the entire treatment course can be used for treatment assessment and outcome analysis, which will be essential to the design and execution of any retrospective and prospective clinical trials utilizing advanced RT treatment techniques.

References


Salama, J. K., Mundt, A. J., Roeske, J., and Mehta, N. 2006. Preliminary outcome and toxicity report of extended-field,


11.1 Introduction: Rationale for Monte Carlo-Based Treatment Planning Systems for Electron Beams ..................................................................................................................155

11.2 Research and Commercial Monte Carlo Treatment Planning Systems........................................157

11.3 Commissioning of MC-Based TPS..............................................................................................159

11.4 Issues Arising from the Clinical Implementation of MC Dose Calculations ................160

11.5 Summary.....................................................................................................................................164

References.............................................................................................................................................164

11.1 Introduction: Rationale for Monte Carlo-Based Treatment Planning Systems for Electron Beams

This chapter covers the practical aspects of Monte Carlo (MC) implementation, the challenges encountered, the main application areas, comparison with conventional algorithms, and so on. Ideally, the discussion will include implementation of commercial packages and research MC packages.

The goal and at the same time the greatest challenge of radiation therapy is to kill the tumor and to spare healthy tissues that do not require treatment. Lots of effort have been devoted to technological developments that allow achieving that goal and improve treatment outcomes. As a result, modern radiation therapy is a complex process that consists of multiple steps, with each step containing inherent uncertainties and assumptions. It has been stated that the uncertainty of dose delivered to the patient should be less than 5% (ICRU-29 1978; Brahme 1984; Papanikolaou et al. 2004). This in turn requires that the uncertainty of dose calculation be less than 2% (Papanikolaou et al. 2004). Most of the commercial treatment planning systems (TPS) cannot provide this level of accuracy for all clinically encountered situations. This has been especially true for electron beams for which traditional dose calculation algorithms fail in many cases (Hogstrom et al. 1981, 1984, 1989; Hogstrom and Almond 1983; Cygler et al. 1987; Ding et al. 2005). In principle, MC gives an accurate answer to within the statistical uncertainty, as there are no significant approximations (except in beam models) involved in the calculations, no approximate scaling of dose kernels is needed and the electron transport is fully modeled. Both treatment machines and patient geometries can be modeled accurately. All types of inhomogeneities are properly handled in the calculations. It has been well documented that MC calculations can be very accurate (Ma et al. 1997; Zhang et al. 1999; Reynaert et al. 2007). In spite of that, for many years, MC calculations were available only for research purposes and used in more advanced departments for special projects or as a quality assurance tool for commercial non-MC systems. This has been mostly due to the long calculation times for MC algorithms. That problem is now overcome with the arrival of modern fast computers and efficient calculation algorithms. Finally, at the dawn of the twenty-first century, state-of-the-art MC dose calculations for electron beams were implemented in commercial TPS. Such systems are making their way into the cancer clinics and are becoming a routine tool used for radiotherapy patients. This is truly a major breakthrough in radiotherapy, since no other existing algorithm can calculate the dose from electron beams.
accurately for all treatment situations. There have been several papers describing commissioning and clinical implementation of MC-based TPS for electron beams (Cygler et al. 2004; Ding et al. 2005, 2006; Popple et al. 2006; Fragoso et al. 2008; Edimo et al. 2009; Ali et al. 2011). Recognizing that MC-based TPS are becoming a reality, the American Association of the Physicists in Medicine, AAPM, formed a special task group, TG-105, to address issues related to clinical implementation of such systems. The TG-105 report provides a useful framework for commissioning of MC-based TPS (Chetty et al. 2007).

11.1.1 Advantages of MC versus Pencil Beam Algorithm

Until recently, the commercial TPS for electron beams were based on pencil beam (PB) algorithms. Although at the time of its implementation, the PB approach was a big step forward, it has been known to have very serious limitations. PB-based planning system cannot really handle monitor unit (MU) calculations for arbitrary source-to-surface distance (SSD) values when only one machine of a single SSD is configured in the system. This is due to the fact, that unlike in the case of high-energy photon beams, the electron beam source is not a point, but rather an extended source. The virtual position of this source in the accelerator head depends on beam energy and field size and so the simple inverse square law, ISL, applicable in case of point source does not work for electron beams (Khan et al. 1991; Cygler et al. 1997). To handle monitor unit calculations for extended SSDs in PB systems, one has to configure several virtual machines, one for each SSD. This is not the case in MC treatment planning systems. In such systems, users need to install just a single virtual machine of a standard SSD for each beam energy to achieve monitor units and dose distributions within clinically acceptable accuracy not only at a standard, but also at any extended SSD (>110 cm) (Cygler et al., 2004; Ding et al., 2006).

In addition, the dose distributions in heterogeneous media have large errors for the complex geometries, frequently encountered in human anatomy. This has been well documented in the literature (Cygler et al. 1987; Hogstrom et al. 1989; Starkschall et al. 1991; Shiu and Hogstrom 1991; Keall and Hoban 1996; Ding et al. 2005; Fragoso et al. 2008). Figure 11.1 shows one of the most striking examples of the failure of PB algorithm (Ding et al. 2005). It is the “trachea and spine” phantom used in electron beam commissioning by Cygler et al. (1987) and Ding et al. (2005, 2006). The phantom consists of overlying low- and high-density heterogeneities, as frequently encountered in the human body. In a solid water slab of overall thickness of 6.2 cm, there is an air pipe of 2.5 cm diameter and 10 cm length and about half centimeter below it, there are four 1 cm thick bone cylinders of 2.5 cm diameter each. Figure 11.1 presents the dose profiles measured and calculated at various depths behind the “trachea and spine” phantom for a 9 MeV beam. As one can see, the PB cannot even remotely resolve the shape of the hot and cold spots for this complex geometry, while MC can do it very well. The MC-measured disagreement that we see here in the peaks is of about 4% and can be attributed to the fact that the measurement resolution was 1 mm and the MC calculation voxel was 3.9 mm. The measurement and calculation resolutions could not be the same, because in this software (Nucletron), users have no control over the calculation voxel size. For 2 mm calculation voxel size, one sees excellent agreement with the measured data (Ding et al. 2005, 2006). In addition to their limited accuracy, PB algorithms require relatively long calculation times for multislice
anatomies. For example, on a standard Eclipse TPS work station (with 8 Xeon processors 2.8 MHz), the MC calculations (with 2% statistical uncertainty) took 4 min whereas the PB calculations took 21 min for the same single beam of 9 MeV electron beam and a $10 \times 10 \text{cm}^2$ field size. The calculation speed gain for MC calculations is partially due to more efficient algorithms and to employing multiple processors of the workstation (Eclipse). The PB algorithm calculations for a single beam on Eclipse utilize a single processor.

11.2 Research and Commercial Monte Carlo Treatment Planning Systems

11.2.1 Meeting the Challenges: The OMEGA Project

The MC method is regarded as the “gold standard” in dose calculation, since it is capable of accurately predicting the dose distributions under almost all circumstances. It has played a significant role in electron treatment planning because it provides accurate dose predictions even in complex three-dimensional heterogeneous geometries, where the model-based dose calculations algorithms, such as PB, have shown significant limitations. Recognizing the requirement of accurate patient treatment planning, in 1990 the NIH funded the OMEGA research project. It was a collaborative project between National Research Council of Canada and the University of Wisconsin to develop a fully 3D electron beam TPS based on MC simulation techniques to calculate the dose to patients. The result of this project was a general purpose MC code, BEAM (Rogers et al. 1995), to simulate radiation beams from radiotherapy machines, including high-energy photon and electron beams, Co-60 beams and kilovoltage units. The BEAM code is currently used worldwide for research. It is also employed to generate and verify beam models used in commercial TPS.

11.2.2 Methods to Speed Up the MC Calculations

Computing time used to be considered a major drawback of the MC technique. In order to use MC in routine clinical treatment planning, special methods were developed to speed up MC calculations for these applications.

One of such methods, called “a voxel based electron beam Monte Carlo algorithm,” introduced some simplifications and approximations into the transport algorithm (Fippel et al. 1997; Kawrakow 1997, 2001). These methods have been implemented into commercial TPS for electron dose calculations (Nucletron and Elekta Software) (Cygler et al. 2004; Ali et al. 2011).

The other technique developed to speed up the electron beam calculations is called the macro Monte Carlo method (MMC) (Neuenschwander and Born 1992; Neuenschwander et al. 1995). The key features of the MMC transport model are: (a) Conventional Monte Carlo simulations of electron transport are performed in well-defined local geometries, namely, spheres (“kugels”). The result of these calculations is a library of probability distribution functions (PDFs) of particles emerging from the “kugels.” The PDFs used in a commercial TPS (Eclipse) were generated in extensive precalculations by employing the EGSnrc (Kawrakow and Rogers, 2002) software to simulate the transport of vertically incident electrons of variable energies through macroscopic spheres of various sizes and five materials (air, lung, water, Lucite, and solid bone). They were calculated only once for a variety of clinically relevant energies and the above-mentioned five materials; (b) patient-specific MC calculations are performed in a global geometry. Electrons are transported through the patient in macroscopic steps based on the PDFs generated in the above described local calculations. As already mentioned, this method has been implemented in a commercial TPS by Varian (Ding et al. 2006; Popple et al. 2006).

Another method to speed up MC dose calculations is to use source (beam) models instead of direct particle transport through the machine head.

11.2.3 EGS-Based MC Research Package

MC particle transport techniques were first introduced in the 1940s and have since evolved into many different areas of applications including radiotherapy physics. In the last several decades, the MC technique has become ubiquitous in medical physics (Rogers 2006). There are many MC codes available for the applications in radiotherapy (Nelson et al. 1985; Seltzer 1991; Sempau et al. 1997; Briesmeister 2000; Kawrakow 2000; GEANT 2006). BEAM/EGS4, a general purpose MC user code for simulation of radiotherapy beams from treatment units was made available in 1995 (Rogers et al. 1995). This code makes it easier to simulate radiotherapy units and has been widely used for the study of photon and electron beams from accelerators. Owing to the easy interface for using the BEAM/EGS4 research package, the BEAM code has been used extensively to characterize therapy beams, including megavoltage electron and photon beams from linear accelerators and kilovoltage photon beams from x-ray units (Ding 1995; Ding and Rogers 1995; Zhang et al. 1998; Zhang et al. 1999; Francescon et al. 2000, 2009; Sheikh-Bagheri and Rogers 2002a,b; Keall et al. 2003; Ding et al. 2006; Jarry et al. 2006; Bazalova and Verhaegen 2007; Pimpinella et al. 2007; Francescon et al. 2008; Bazalova et al. 2009; Iaccarino et al. 2011; Pasciuti et al. 2011). Not only is the BEAM/EGS4 (now BEAMnrc/EGSnrc) research package capable of simulating the treatment beams, but also uses the simulated realistic beams to calculate dose to a patient using the information from CT images. MC research packages have also been used in validating the accuracy of dose calculations of commercial TPS (Ma et al. 2000; Wieslander and Knoos 2007; Ali et al. 2011).

11.2.4 Commercial MC TPS

MC calculations of dose distribution in a patient can be divided into three steps (Chetty et al. 2007):

1. Particle transport through the top of machine head (patient-independent step).
2. Particle transport through the patient-specific part of the machine.
3. Dose calculations in the patient.

In research, for particle transport through the machine head, BEAM or BEAMnrc are frequently used. BEAM generates a particle phase-space file, which is used as the input file to steps 2 and 3, dose calculations in a patient. In step 2, codes such as VMC, XVMC, VMC++ are used.

Direct simulation of particles in the machine head is not used in commercial TPS. The main reason is that simulations of the direct particle transport through the machine head strongly depend on knowing accurately the machine head details (materials, dimensions, etc.), which may not be easily available to users (Schreiber and Faddegon 2005).

So, typically commercial MC-based TPS use source models, which describe the beams with several fitting parameters derived from the fully simulated phase-space file and/or measured dose profiles.

Presently, there are three commercial MC-based TPS available: Nucletron Oncentra MasterPlan (VMC++), Varian Eclipse (eMC), and Elekta Software XiO® (XVMC).

11.2.4.1 Nucletron VMC++

In 2002, Nuclotron* was the first in the world to release a commercial MC-based TPS for electron beams. The MC option is available in the Dose Calculation Module, DCM, of Theraplan Plus and Oncentra MasterPlan. The beam transport is divided into two components: treatment independent and treatment dependent (Oncentra MasterPlan—Physics Reference Manual). The treatment-independent component, called phase-space engine, handles the particle transport through the linac head down to the lowest patient-independent collimation level. The patient-dependent component, called the dose engine, is based on the VMC++ algorithm (Kawrakow 2000). The accelerator beam model in Nucletron TPS was developed by Traneus and colleagues (Traneus et al. 2001). It consists of the coupled multisource beam model. It is based on five parameters and requires a limited set of clinic-measured data to configure the beam model. This beam model is flexible enough to support current collimation types such as applicators with optional inserts, variable trimmer machines, and potential multileaf collimator (MLC), applications in the future. The modeling of a particular beam starts upstream of the uppermost changeable/movable collimating element with a source phase space (SPS). The SPS is propagated through the treatment head to an exit phase space (EPS) plane located at the lowest collimating element thereby defining the interface between the phase space and the dose engines, see Figure 11.2.

For a treatment unit with fixed applicators with optional patient-specific cutouts, the block collimators and possible multileaf collimator (MLC), applications in the future. The modeling of a particular beam starts upstream of the uppermost changeable/movable collimating element with a source phase space (SPS). The SPS is propagated through the treatment head to an exit phase space (EPS) plane located at the lowest collimating element thereby defining the interface between the phase space and the dose engines, see Figure 11.2.

For a treatment unit with fixed applicators with optional patient-specific cutouts, the block collimators and possible multileaf collimator (MLC), applications in the future. The modeling of a particular beam starts upstream of the uppermost changeable/movable collimating element with a source phase space (SPS). The SPS is propagated through the treatment head to an exit phase space (EPS) plane located at the lowest collimating element thereby defining the interface between the phase space and the dose engines, see Figure 11.2.

For a treatment unit with fixed applicators with optional patient-specific cutouts, the block collimators and possible multileaf collimator (MLC), applications in the future. The modeling of a particular beam starts upstream of the uppermost changeable/movable collimating element with a source phase space (SPS). The SPS is propagated through the treatment head to an exit phase space (EPS) plane located at the lowest collimating element thereby defining the interface between the phase space and the dose engines, see Figure 11.2.

For a treatment unit with fixed applicators with optional patient-specific cutouts, the block collimators and possible multileaf collimator (MLC), applications in the future. The modeling of a particular beam starts upstream of the uppermost changeable/movable collimating element with a source phase space (SPS). The SPS is propagated through the treatment head to an exit phase space (EPS) plane located at the lowest collimating element thereby defining the interface between the phase space and the dose engines, see Figure 11.2.

For a treatment unit with fixed applicators with optional patient-specific cutouts, the block collimators and possible multileaf collimator (MLC), applications in the future. The modeling of a particular beam starts upstream of the uppermost changeable/movable collimating element with a source phase space (SPS). The SPS is propagated through the treatment head to an exit phase space (EPS) plane located at the lowest collimating element thereby defining the interface between the phase space and the dose engines, see Figure 11.2.

By sampling from a parameterized description of this EPS. In this way, the transport through the treatment head does not have to be performed for every dose calculation. For more general treatments that are not using fixed collimation setups (e.g., MLC fields) the dose calculation can be driven by sampling directly from the SPS.

To by-pass time-consuming in-collimator transport during dose planning, scattering effects in the applicator are accounted for by using precalculated collimator scatter kernels, describing the coupled angular and energy probability distributions of indirect electrons coming from applicator elements.

Therefore, for the Nucletron planning system, the accelerator model provides the EPS file, which serves as the input file for the in-phantom/patient calculations performed with VMC++ (Kawrakow et al. 1996; Kawrakow 1997, 2001).

The code can handle fixed applicators with optional, arbitrary inserts of any shape. It also can handle the variable size fields defined by an applicator like DEVA (Siemens). The code always calculates the absolute dose per MU, cGy/MU.

Theraplan Plus and Oncentra planning systems can handle all major accelerators available on the market. The beam modeling is performed by Nucletron.

* It actually was MDS Nordion who later sold the TPS to Nucletron.
The user interface for the Nucletron planning systems does not allow for selection of any calculation parameters, except the number of particle histories/cm². The larger this number, the lower the statistical uncertainty in the dose distribution. The user has no control over the calculation’s voxel size, which is automatically assigned by the system. The voxel size assignment is based on the CT anatomy volume and the total number of allowed voxels being 800,000. So for smaller CT volumes, one achieves smaller calculation voxel sizes. The larger overall volumes result in being divided into larger calculation voxels. A database of 21 materials is used in segmentation of patient CT anatomy.

In Theraplan Plus, the user can select to calculate and report dose-to-medium or dose-to-water. Unfortunately, the user does not have this choice in Oncentra, which reports dose to water.

### 11.2.4.2 Varian eMC

The electron Monte Carlo (eMC) algorithm employed by Eclipse is an implementation of the MMC method (Neuenschwander et al. 1995) for calculation of dose from high-energy electron beams. The algorithm consists of: (a) an electron beam phase-space model describing the electrons that emerge from the treatment head of the linear accelerator. This model is based on precalculated data for a machine type and configured using measured beam data; (b) electron transport/dose deposition in the patient is calculated using the MMC method. The patient’s anatomy is based on CT images. The CT image volume is first converted into a mass density volume with a user-defined resolution (0.1–0.5 cm), applying appropriate CT-to-mass density conversion factors. The resulting density volume is then scanned for heterogeneities. To each voxel of the density-volume, a sphere index is assigned. A voxel of the density volume is then scanned for heterogeneities. To each voxel of the density-volume, a sphere index is assigned. A voxel of the density volume is considered to be part of a heterogeneous volume if the density ratio of the voxel and its neighbors exceed a limit (typically 1.5). If the densities in both voxels are below a threshold (typically 0.05 g/cm³), the ratio is not evaluated. The material assigned to each sphere depends on the average mass density within the sphere. If the average mass density of a sphere is exactly equal to the mass density of one of the preset materials, the preset material is selected for the sphere. If the mass density of a sphere is between two preset materials, the material is randomly selected for these two materials each time a particle enters the sphere. The probability for a material to be selected is proportional to the closeness of a sphere’s average mass density to the mass density of the material. There are five preset different materials (air, lung, water, Lucite, and bone) in the MMC calculation database (The Reference Guide for Eclipse Algorithms Version 6.5 (P/N B401653R01M). For example, if the average density within a sphere is 1.12, there is a 12/19 chance that the material is selected to be Lucite and a 7/19 chance that is selected to be water. Unlike conventional TPS, the reported doses are the doses calculated to the material (Ding et al. 2006) and not the dose-to-water of different densities. However, the user can define other calculation parameters, such as number of particle histories or required statistical uncertainty of calculation, voxel size, or whether to apply isodose smoothing.

### 11.2.4.3 Elekta Software XiO eMC

The XiO eMC module is based on XVMC (Kawrakow et al. 1996; Fippel 1999; Fippel et al. 1999). The beam model has been developed by Elekta Software and it consists of a weighted combination of particle sources (direct and indirect electrons and photons). Beam modeling based on user measurements is performed by Elekta Software. As mentioned above, the dose calculation in the phantom/patient is based on XVMC. The user interface in XiO allows for full control over the dose calculation parameters. The user can define calculation voxel dimensions and can control the statistical uncertainty in the dose distributions by setting either the maximum number of particle histories or the goal MRSU (the mean relative statistical uncertainty) defined as

$$\text{MRSU} = \sqrt{\frac{1}{N} \sum_{i=1}^{N} \left( \frac{\sigma_i}{D_i} \right)^2}$$

where $D_i$ is the dose in $i$th voxel, $\sigma_i$ is the statistical uncertainty of dose for the $i$th voxel. The sum is computed over all $N$ voxels in which the dose is greater than or equal to some fraction, $\alpha$, of the maximum dose $D_{\text{max}}$. It is commonly accepted that for tumor dose calculations $\alpha = 50\%$. However, for accurate calculation of dose to organs at risk, $\alpha$ should be set to a much lower value, which will of course increase the calculation time. XiO eMC allows the user to calculate and report either dose-to-water or dose-to-medium, which makes it compliant with TG-105 recommendations (Chetty et al. 2007). XiO can handle calculations for linacs from all major vendors.

### 11.3 Commissioning of MC-Based TPS

Commissioning of MC-based TPS can be divided into three logical steps:

1. Collection of data to allow beam modeling, as outlined in the user manual.
2. Verification of the beam model.
3. Verification of dose calculations in heterogeneous media and under typical clinical conditions.

#### 11.3.1 Measurements Required for Beam Characterization

To create a beam model, the user has to provide the following information about the treatment unit to the treatment planning vendor:

- Position and thickness of jaw collimators and MLC.
- For each applicator scraper layer: thickness, position, shape (perimeter and edge), and composition.
- For inserts: thickness, shape, and composition.

The open-field measurements (without the applicator, with collimator jaws wide open) consist of:

- Depth–dose curves in water at SSD = 100 cm.
The applicator measurements for each energy/applicator combination are:

- Depth–dose curves in water at SSD = 100 cm.
- Absolute doses (reference dose), expressed in cGy/MU, at a specified point on the depth–dose curve.

The reference dose (in cGy/MU) for the beam defined by an applicator is normally greater than the reference dose of the open beam (no applicator) due to the applicator scatter.

The electron applicator field size is entered in the beam configuration task. The photon jaw settings are set corresponding to open beam (no applicator) due to the applicator scatter.

11.3.2 Additional Beam Data Measurements for Commissioning MC-Based TPS

To commission a MC-based TPS, additional input data related to the beam characteristics may be needed. For example, to configure the eMC algorithm in the Varian Eclipse system, measurements are required for all open-field and applicator combinations for each beam energy. In addition, the dose profiles in air at a specified source-to-detector distance equal to 95 cm are also required (Ding et al. 2006). In the MC-based TPS, the user typically needs to measure for each beam energy dose profiles in air and water, both with and without the applicators (Cygler et al. 2004; Ding et al. 2006). The open-field (beam without electron applicator) beam data are necessary in order for the TPS system to accurately model the incident electron beams for dose calculations. In addition, the in-air profiles provide incident electron beam information (without the effect of an applicator) needed to configure the beam (Cygler et al. 2004; Ding et al. 2006).

The beam modeling procedure in XiO is based on in-water measurements only. It requires for each energy and applicator depth–dose curves and profiles at 7 depths for both an SSD of 100 cm and for an extended SSD commonly used in the clinic. The same set of scans is also required for a 15 × 15 cm² applicator with a 5 × 5 cm² cutout in place. In addition, depth–dose curves and profiles at 7 depths are also required for SSD = 100 cm with no applicator present and the jaws set to their maximum. The user has to also provide point dose values at a reference depth (typically $d_{\text{max}}$) for each field size and SSD listed above.

11.3.3 Beam Measurements Required for In-Phantom/Patient Dose Calculation Verification

Good clinical practice requires careful commissioning of a TPS. During the commissioning, the user should verify the accuracy of dose calculations of a TPS before such a system is put into clinical use. This can be done by performing calculations and comparing the results with the data measured in homogeneous and some heterogeneous phantoms.

11.3.3.1 Homogeneous Phantom: Dose Profiles, MU Calculations at Various SSDs

Water is normally the choice of homogeneous phantom because the dose distributions can be obtained through depth–dose curve or dose profile scanning measurements by using commercially available scanning systems with an ionization chamber, diode or other type of detector. The measurements of depth–dose curves and dose profiles at several different SSD that are clinically relevant (i.e., SSD = 100 cm, 115 cm) are necessary not only for beam modeling but also for verification of dose calculations at standard and extended SSDs. The validation should also include the accuracy of MU calculations at standard and extended SSDs for open applicators and for a variety of cutout-defined field sizes for each specified electron applicator. One of the important strengths of the MC-based electron TPS is that it is capable of accurately calculating both dose distributions and MUs not only at a standard SSD but also at any extended SSD (Cygler et al. 2004; Ding et al. 2006). This is a significant advantage because conventional electron dose calculation algorithms are not capable of accurately predicting either the dose distributions or MUs at an extended SSD when the beam is commissioned only at the standard SSD (Ding et al. 1999, 2005).

11.3.3.2 Heterogeneous Phantoms

In addition to the validation of dose calculation in homogeneous phantoms, some tests should be carried out for heterogeneous phantoms before clinical implementation of the TPS. Such tests should include heterogeneous phantoms of various complexities such as 1D-type or slab, 2D-type, and more complex 3D-type geometries. The density of the heterogeneous materials must include high-(bone) and low-(lung) density materials that are clinically relevant in the patient treatment planning calculations. The validation involves the comparison of doses calculated by the TPS and the measured ones. Studies by Cygler et al. (2004) and Ding et al. (2006) showed some possible methods of evaluating the accuracy of a commercial electron TPS in a complex phantom containing 3D-type heterogeneities.

11.4 Issues Arising from the Clinical Implementation of MC Dose Calculations

11.4.1 Calculation Normalization, Dose Prescription, and Isodose Lines

As part of the commissioning of an electron TPS for clinical use, we also validated the accuracy of this system to calculate output factors for various cutout sizes for arbitrary SSDs for all available beam energies. A clinically acceptable agreement between measured and calculated output factors should be obtained. This includes both...
FIGURE 11.3  (a) Effect of using different smoothing techniques with the same calculation voxel sizes on the dose distributions. (b) Effect of using different calculation voxel sizes and smoothing techniques on the depth–dose distributions. (c) Effect of statistical uncertainties on dose distributions in the MC-based TPS. (Reproduced from Ding et al. 2006. *Phys Med Biol* 51(11):2781–99. With permission.)
open applicator and cutout-defined fields. When a calculated dose distribution is not normalized appropriately in a TPS, user applied renormalization may be necessary in order to obtain accurate MUs. It has been shown that MC methods available in commercial TPS are capable of calculating output factors accurately even at extended SSDs. This is a significant improvement over the PB algorithm implemented in CADPLAN (Ding et al. 2005).

### 11.4.2 Statistical Uncertainty, Smoothing and Calculation Voxel Size

It is noteworthy that any MC calculation includes a statistical uncertainty. It is not easy to have a rigorous method to obtain an accurate statistical uncertainty of the calculated dose distribution. Therefore, if a target prescription dose is based on a point dose calculation, the statistical uncertainty may have a significant effect on the dose prescription. The tools provided for smoothing dose distributions in some commercial TPS do not discriminate between real dose gradients and statistical variations. When applying the smoothing tools or using large calculation voxels to reduce the computation time, caution must be exercised or the resultant details of the calculated dose distributions may not be accurate, especially in regions of high-dose gradients (Ding et al. 2006a,b). On the other hand, the statistical uncertainty can be misinterpreted as dose variation when the real statistical uncertainty of the calculation is not well known. Figure 11.3 illustrates the effect of using different calculation voxel sizes, smoothing techniques, and statistical uncertainties on dose distributions in the MC-based TPS.

### 11.4.3 Dose-to-Medium versus Dose-to-Water Calculations

One of the issues arising when using MC-based TPS is “dose-to-water vs. dose-to-medium” calculations (Liu 2002; Keall 2002; Ding et al. 2006; Gardner et al. 2007). The two are conceptually different and which one is used has a significant impact on the reported doses to different organs, such as bone and lung. In conventional TPS, CT numbers are converted to electron densities of water-like material before calculations are carried out. Unlike conventional TPS, MC inherently calculates the dose to medium. There can be deviations from unity (>10%) in reported doses when dose to water as opposed to dose to medium is calculated. This is due to the fact that the stopping-power ratios of water/medium may not be unity for electron beams. There could be significant differences. The study by Ding et al. (2006) provided some examples of such situations. Their investigation found significant differences for lung and bone, respectively, when “dose-to-medium” is reported instead of “dose-to-water” as done in the conventional planning systems. In addition, the magnitude of this difference is not constant but depends on the beam energy, calculation depth, and medium. This variation is due in part to the fact that water/medium stopping-power ratios are strongly dependent on beam energy compared to photon beams. Figures 11.4 and 11.5 illustrate the magnitude of this difference for hard bone. This large variation due only to the method of dose reporting may have significant clinical implications. As MC-based TPS begin to enter routine clinical practice, it is imperative that a consistent approach to dose reporting is used. Only then can the prescription dose and treatment outcomes can be compared in a meaningful way between different centers.

### 11.4.4 Examples of CT-Based Dose Calculations

Figure 11.6 shows a calculated dose distribution for a tumor near the left ear. It can be seen that the MC calculated dose distributions show a significant hot and cold regions due to the presence of low- and high-density media whereas the PB dose calculation algorithm is not able to accurately predict the dose changes caused by the inhomogeneity. The ability to accurately predict hot and cold regions in patient treatment planning is essential to provide the best possible choice for treatment options.

### 11.4.5 Typical Calculation Times

MC dose calculation times depend on the computer hardware, dose calculation algorithm, beam energy, field size, and

---

**FIGURE 11.4** Illustration of water-to-medium stopping power ratios as a function of depth for an 18 MeV electron beam in water. It is seen that the magnitude of differences between (a) dose-to-bone or (b) dose-to-lung and dose-to-water can be up to 12% or 4%, respectively. (Reproduced from Ding et al. 2006. Phys Med Biol 51(11):2781–99. With permission.)
**FIGURE 11.5** (See color insert.) Clinical example of differences between dose-to-medium and dose-to-water calculations for 6 MeV beam (XiO eMC).

**FIGURE 11.6** Calculated dose distributions using MC algorithm (a) and PB algorithm (b) in a commercial TPS. It is seen that PB dose calculation algorithm is not able to accurately predict dose changes caused by the inhomogeneity. The dose profiles start at point A and end at point B as shown in the axial CT images.
calculation parameters. The parameters that affect the speed of dose calculation are the number of particle histories or required statistical uncertainty and the voxel size. Users of a commercial TPS do not always have control over all parameters that affect the speed of dose calculations. Elekta Software is the vendor that allows users to set the largest range of calculation parameters. The proper choice of parameters depends on an anatomical site and required statistical uncertainty. Figure 11.7 shows typical calculation times for the trachea and spine phantom using a clinical XiO (version 4.51) workstation (Linux OS, 8 CPU 3 GHz each, 8.29 GB of RAM) for two beam energies and two voxel sizes as a function of MRSU. Typical calculation times for other planning systems can be found in Cygler et al. (2004), Ding et al. (2005), Chetty et al. (2007), and Fragoso et al. (2008).

11.5 Summary

Commercial MC-based TPS are becoming more common in the clinics. They provide significant improvement in dose calculation accuracy over the PB algorithms. It has been shown that in homogeneous phantoms both PB and MC systems are able to predict dose distribution accurately for standard SSDs. However, PB calculations are not accurate in a homogeneous water phantom for extended SSDs. For such cases, MC provides significant improvements. Only one virtual electron beam for each energy is used for accurate dose and MU calculations at any arbitrary SSD as demonstrated in the study by Cygler et al. (2004) and Ding et al. (2005). In heterogeneous phantoms, MC dose calculation method offers superior accuracy as compared to PB algorithm, especially where small 3D heterogeneities or overlying low-(air) and high-(bone) density materials are present. MC results have generally much better agreement with measurements, especially in high-dose gradient regions caused by the perturbation of adjacent 3D inhomogeneities. The overall results reported in the literature clearly underscore the potential of MC to become the method of choice in the electron beam treatment planning compared to PB algorithms. However, it is important for the user to understand the MC implementation in a particular commercial TPS in order to select appropriate calculation settings for a given clinical situation. It is also worth noting that there is no absolute guarantee that the result from the MC calculation is always accurate.

References

Ding, G. X. 1995. An investigation of radiotherapy electron beams using Monte Carlo techniques, PhD thesis, Department of Physics, Carleton University, Ottawa Carleton Institute for Physics. PhD, Ottawa, Canada.


12.1 Introduction

The principal aim of radiotherapy is to give a tumoricidal dose to the cancer-bearing tissue and to minimize the dose to normal healthy tissue. Given the dose response function of the tumor and the healthy tissue, the requirement stated in the ICRU report 50 (International Commission on Radiation Units and Measurements Report 50, 1993) is to apply the dose to the tumor within $\pm 5\%$ and $7\%$ of the prescribed dose. A detailed analysis of uncertainties associated with radiation treatment shows that $3\%$ accuracy is required in the dose calculation to yield $\pm 5\%$ accuracy in the dose delivered to the patient (International Commission on Radiation Units and Measurements Report 24, 1976; Brahme, 1984; Dutreix, 1984; Van Dyk et al., 1993). Some studies have even concluded that for certain types of tumors the uncertainty in dose delivery should be smaller than $3.5\%$ (Brahme, 1984; Mijnheer et al., 1987, 1989), which in turn means that the clinically implemented dose calculation algorithm should be accurate within $\pm 2\%$. Thus, accurate dose calculation algorithms for treatment planning systems are of critical importance in radiation therapy.

Within the last two decades, huge developments have been seen in the field of dose calculation algorithms. The early clinical treatment planning systems used correction-based algorithms (Mackie et al., 1996), whereas nowadays mostly model-based algorithms such as convolution or convolution/superposition algorithms are used. Algorithms using the Monte Carlo (MC) method can also be considered as model based. Although over the last few years many applications of MC techniques have been published in the field of medical physics, the clinical impact of MC calculated patient dose distributions for photon beams still remains unclear.

In contrast to the other common techniques, the MC method starts from first principles and tracks individual particle histories; thus, it takes into account the transport of secondary particles. For a long time, a major drawback of using MC methods for photon beam treatment planning in clinical routine was the long calculation time needed to achieve a dose distribution with a reasonable statistical uncertainty. However, recent advances in MC patient dose calculation algorithms coupled with increasing computer processing speed has made MC patient dose calculation speed acceptable for the radiotherapy clinic. In principle, the MC method produces accurate results in regions of tissue heterogeneities such as lung and surface irregularities, thus providing the most accurate method for the simulation of patient treatment dose distributions, especially for complex techniques such as intensity modulated radiotherapy (IMRT) or intensity-modulated arc therapy (Andrelo, 1991; DeMarco et al., 1998; Wang et al., 1998; Arnfield et al., 2000; Bush et al., 2008; Oliver et al., 2008; Sarkar et al., 2008; Teke et al., 2010; Fix et al., 2011). The National Cancer Institute recognizes the need for research and development on MC techniques in radiation therapy and it is anticipated that MC will become a necessary dose calculation tool (Fraas et al., 2003). However, ultimately the accuracy of the treatment planning dose calculation algorithm depends strongly on the implementation and the accuracy of the input
data. For instance, the accuracy depends on the quality of the anatomical information of the patient, since this affects the irradiating geometry as well as the tissue cross sections. Another prerequisite is the accurate information about the radiation beam incident on the patient. In fact, given the availability of fast radiotherapy-specific MC codes, the major limitation to the widespread implementation of MC dose calculation algorithms is the lack of a general, accurate, and user-specific scalable source model of the accelerator radiation source. More precisely, a user with an arbitrary linear accelerator should be able to commission the source model so that the MC dose calculation algorithm meets predefined accuracy requirements compared to measurements prior to using the algorithm for patient dose calculation, for example, 2% or 2 mm.

12.2 Requirements for Clinical MC Treatment Planning

A clinically usable MC treatment planning system for photon beams is more than just an MC dose calculation algorithm together with a beam model to describe the radiation beam. Additionally, it is important to also have the capability, that is, a tool or a platform, for beam setup, dose display, and dose evaluation. Whereas these platforms are available in commercial treatment planning systems, they are in general missing in the preclinical or research MC dose calculation packages. If MC dose calculation is used only in some special rare cases, this might be acceptable. However, if MC treatment planning should be used on a large scale, automation of the MC dose calculation is needed. Consequently, some research systems have also been extended to provide capabilities for beam setup, dose display, and dose evaluation. The following sections will cover the requirements for clinical MC treatment planning for photon beams.

12.2.1 Beam Setup Capability

For the commercial treatment planning systems, this is the same as for any other dose calculation algorithm. But research systems, if intended for clinical use, also provide such a platform. Most commonly, this has been realized by interfacing the external MC dose calculation with an already existing commercial treatment planning system automatically (Fix et al., 2007; Siebers et al., 2000), by another interfacing program (Ma et al., 2002), or by a DICOM-RT interface (Alexander et al., 2007).

12.2.2 Beam Model

Certainly, an accurate characterization of the radiation beam exiting the treatment head is a prerequisite for accurate dose calculations in the patient. For MC photon treatment planning, several beam models have been developed and investigated. They are either measurement or MC based (Verhaegen and Seuntjens, 2003). The measurement-based beam models use analytical models for which the parameters are determined from the measurements (Ahnesjo et al., 1992, 1995; Ahnesjo, 1994; Jiang et al., 2001; Fippel et al., 2003). The MC-based beam models can be further divided into those using full MC (Hasenbalg et al., 2008) or phase space files (Rogers et al., 1995; Sheikh-Bagheri and Rogers, 2002a,b) as input and those using histogram-based beam models (Ma, 1998; Schach von Wittenau et al., 1999; Deng et al., 2000; Fix et al., 2004). These MC-based beam models usually model the primary beam, that is, the part of the accelerator head that covers the patient-independent part.

Besides the characterization of the primary beam, the beam model also has to accurately model the clinically used patient-specific beam modifiers. For photon beams, these are blocks, hard wedges, dynamic wedges, multileaf collimators (MLCs), and so on (Magaddino et al., 2011). For the radiation transport through these beam modifiers, a number of transport parameter settings and dedicated variance reduction methods can be applied depending on the treatment case (Schmidhalter et al., 2010). In general, the research systems provide a higher degree of flexibility compared to commercial treatment planning systems with respect to transport options and also on how to use the beam modifiers. For example, the reduction in the MLC transmission for moving jaws in IMRT has been investigated (Schmidhalter et al., 2007). Those settings need to be considered very carefully before using them for clinical MC dose calculations.

Since errors in the beam model propagate through all the following processes for the dose calculation, it is very important to extensively verify the performance of the beam model during the commissioning and validation process.

12.2.3 Patient Model

Another important issue affecting the accuracy of the MC calculated dose distribution in the patient is the anatomical patient representation, which is the basis for the geometrical and interaction data specification used for the dose calculation. Patient computed tomography (CT) scans are used to extract the geometrical information as well as the patient-specific tissue characterization leading to the interaction data for the dose calculation within the patient. Therefore, image artifacts lead to inaccurate patient representation and consequently to inaccurate dose distributions. For non-MC dose calculation algorithms, CT conversion curves from Hounsfield values to electron density or physical density are used. However, for MC algorithms, the interaction data also have to be known. Different conversion methods are used starting from a simple relation between mass density and interaction coefficients (Fippel, 1999) to explicit segmentation of the tissue followed by anatomic composition assignment to determine the interaction data (Chetty et al., 2007). The impact of material misassignment on dose distributions is not fully investigated yet and there are only a few studies in the literature (Verhaegen and Devic, 2005; Vanderstraeten et al., 2007; Ottoisson and Behrens, 2011). Detailed information about this issue can be found in Chapter 8 of this book. Another issue with CT images is the grid resampling, since typically a different voxel size is used for the dose calculation than provided by the CT images. The applied
interpolation method can introduce errors, which in turn lead to misassignments in the materials and finally in the interaction data. A recent study by Volken et al. (2008) investigating different interpolation algorithms shows that an integral conservative Hermitian curve interpolation improves the interpolation accuracy compared with typically used linear or cubic interpolation functions. Finally, it is worth mentioning that changes occurring between the image acquisition and the dose delivery need to be taken into account. This could be changes to the patient, for example, weight loss, accessories only used during CT image acquisition, or different couches used for the CT scan and radiation delivery. Consequently, it might be necessary to remove that information from the CT images or overwrite the information with more accurate data before using them for dose calculations.

12.2.4 Dose Calculation

The MC dose calculation algorithm needs to be interfaced with the beam model. In principle, there exist several methods. One approach is to store phase space files as output of the beam model and use these phase space files afterwards as input for the dose calculation, that is, the phase space file is used as beam model in this case. Another approach is to pass the particle from the beam model directly, that is, in memory, to the dose calculation algorithm. The latter has the advantage of being faster and that no large phase space files have to be stored on hard discs. The fact that MC dose calculation is dealing with random walks offers the possibility to reuse the same particle coming from the beam model. However, by this method, the statistical uncertainty of the calculated dose distribution is affected, since those particles are no longer independent. Thus, whereas this method saves some computing time and possibly disc space, the statistical uncertainty of the representation of the radiation beam by the beam model reveals some lower limit. One example is the latent variance of a phase space file (Sempau et al., 2001). Alternatively, modifications of the particles from the phase space utilizing existing symmetries (Rogers et al., 1995; Fix et al., 2004; Bush et al., 2007) or small perturbations (Tyagi et al., 2006) have been investigated.

One of the main advantages of using MC dose calculation is that such an algorithm is suitable not only to calculate the dose for static patient situations but also for dynamic situations, like patient motion. Since MC is able to include a time component—in the beam model as well as in the patient model—MC can be used straightforward for dynamic patient dose calculations (Paganetti et al., 2004, 2005; Seco et al., 2008). Recently, new dynamic arc therapy techniques such as RapidArc, IMAT, or VMAT have been introduced in clinical routine. It is worth mentioning that for these applications the dose calculations with MC methods can be performed continuously instead of using a discrete number of situations approximating the dynamic delivery, for example, multiple static beams every 5° to approximate the gantry rotation. It is also worth mentioning that the calculation time for MC dose calculations does not scale with the number of beams if the statistical uncertainty in the target is considered as compared to other conventional algorithms. However, the calculation time might increase to some extent if a certain statistical uncertainty has to be achieved in organs at risk (OARs).

12.2.5 Dose Evaluation Capability

Similar to the beam setup, the dose evaluation can also be performed as for any other dose calculation algorithm; thus, the commercial treatment planning systems are able to use their already existing tools for dose display and dose evaluation. This is true also for those research systems that are automatically interfaced with commercial treatment planning systems or those using DICOM files as interfaces. However, additional important information is needed, which is generally not included in commercial dose display and evaluation tools. Since every MC simulation is associated with a statistical uncertainty, the documentation of this information for the dose evaluation is essential when dealing with MC. Bearing in mind that in the dose evaluation different structures are quantitatively analyzed, it is important to take into account the statistical uncertainty of the dose values within the structure, for example, by displaying the uncertainty distribution. Additionally, it should be stated whether dose to media or dose to water has been calculated independent of which dose calculation algorithm is used. Another possibility to display and evaluate the resulting dose distribution is external viewing and analysis tools, for example, the CERR software development (Deasy et al., 2003).

12.3 Commissioning and Validation

Usually, vendors of commercial treatment planning systems provide a recommendation of the commissioning and validation procedure for their planning systems. Thereby, the commissioning consists of the configuration of the treatment planning system. The commissioning and validation are performed by comparing measured with the corresponding calculated dose distributions for different setups. However, especially for MC implementations, it is not fully explored what the least sufficient set of comparisons is for the commissioning and validation process. Of course this depends strongly on the clinical use foreseen. For example, which type of treatment technique should be covered: 3D conformal, IMRT, stereotactic radiotherapy, and so on. It is further important to determine what kind of beam modifiers have to be provided. Finally, it also depends on the accuracy requested, detector types used for the measurements, and so on. The following sections provide an overview for the commissioning and validation of MC treatment planning systems.

12.3.1 Tolerances and Acceptance Criteria

During the commissioning and acceptance procedure, it is always a difficult task to determine the tolerance and acceptance criteria for the dose comparisons. Several quantities have been used for comparing two dose distributions, for example, dose difference, distance to agreement, gamma index (Low et al., 1998), or slightly
modified gamma index values (Bakai et al., 2003; Jiang et al., 2006; Blanpain and Mercier, 2009). For these quantities, a certain tolerance level or acceptance criteria have to be defined. Often, 2% or 2 mm is chosen. However, if such criteria are required for patient dose calculation, the estimation for the error due to beam modeling may not be larger than 1% or 1 mm (Keall et al., 2003).

It is also questionable to apply the same criteria throughout the comparison, that is, for all field sizes, at every location, and so on. Given the statistical nature of MC, there is a certain probability of dose values with large random errors that might not fulfill the criteria. But there might also be areas in which the accuracy is more important than in others, for example, build-up region or outside the direct radiation field. Thus, different criteria at different locations and setups might be justified. In summary, it is of great importance to carefully choose the tolerance and acceptance criteria ahead of the commissioning and validation procedure and in relation of the clinical usage of the machine.

12.3.2 CT Conversion

As already mentioned in previous sections, it is of great importance to have accurate anatomical and material information. Depending on the MC algorithm, different conversion methods are needed to finally receive the interaction data needed for the MC dose calculation. The according input information for the conversion method used has to be provided to the system by the user. Most MC treatment planning systems, including the research systems, use a mapping of Hounsfield values to mass density and material composition. In the beginning of MC photon treatment planning, often a default CT to material conversion has been used. This conversion is not appropriate if highly accurate dose calculation is required. Incorrect assignment of media and/or density can lead to significant dosimetric errors (see Section 12.2.3). A more accurate CT to material conversion requires more segmentation bins compared with the default scheme (Vanderstraeten et al., 2007) and depends on the CT scanner used. Hence, CT calibration phantom used for the CT scanner in use lead to more accurate patient representations and material specifications. However, the phantom has to be carefully evaluated to determine if it is suitable for such a calibration procedure, for example, Teflon is not an appropriate representation for cortical bone (Verhaegen and Devic, 2005).

12.3.3 Beam Model and Dose Calculation

The specific beam model commissioning depends strongly on the implemented beam model. Most research-based systems have a slightly different implementation and thus an individual beam model commissioning procedure. For beam models based on phase space data, usually the energy and the width of the Gaussian intensity distribution of the initial electron beam are used as parameter. Note that perfect knowledge of the treatment head geometry and material is assumed. Beam models derived from these phase space data usually have settings available for a wide range of different parameterized initial electron beams.

During the commissioning, the beam model created is determined by interpolation. Analytical or measurement-based beam models have a specified set of measurements needed for the commissioning. The scope of these measurements is provided either by the vendors or by the research group.

Beam modifiers typically represent the patient-specific part of the linear accelerator within the beam model. For the radiation transport, mainly either MC transport or transmission filters in multiple layers are used. The parameters that are adjusted during the commissioning phase of the beam model depend on the beam modifier considered. Typically, the density is tuned for the secondary collimators and blocks, whereas additionally, the thickness is adjusted for the hard wedges. For the tuning of the MLC, the gap between the leaves is used as a third parameter. As for the open beam, and for the beam modifiers too, a set of measurement is used to determine the settings of the tuning parameters. Since these parameters are not depending on the photon beam energy, the commissioning could in principle be done for only one beam energy. Alternatively and depending on the clinical demands, the commissioning could determine tuning parameters that provide the best fit for all beam energies on average. The validation, of course, has to be done for all beam energies used clinically.

The validation of the beam model is done by comparing calculated with measured dose distributions. The comparisons include the measurements that have been used during the commissioning process to verify that the treatment planning system is able to reproduce the input dose distributions, that is, a consistency test is performed. For the validation, it is important to cover the dosimetric range for which treatment plans are used clinically. Typically, the following measurements are compared:

- Relative depth dose curves and lateral dose profiles for different field sizes and shapes, including off-axis fields. These measurements are needed for open fields as well as for all beam modifiers considered such as wedged, blocked, dynamic wedged, and static and dynamic MLC-shaped fields. These measurements are performed in water or water equivalent phantoms. Since the primary aim is to validate the beam model and the dose calculation, disturbing side effects due to CT conversion or volume averaging and so on should be avoided or minimized. This could be realized by using digital phantoms.
- Since relative depth dose curves and lateral dose profiles are used above, it is necessary to check output factors for the same situations as well as an absolute dose calibration. Alternatively, all dose curves could be compared in absolute units, for example, cGy or Gy per monitor unit (MU).
- Relative measurements using simple inhomogeneous phantoms.
- Clinical treatment plan comparisons between dose distributions calculated with MC and with other available conventional dose calculation algorithms. This should involve simple cases as well as complex cases. Thereby, the patient is set to water with density 1 g/cm³ or with the density
resulting from the CT conversion scheme. This provides some consistency checks with conventional dose calculation algorithms.

Some remarks with respect to the commissioning and validation:

- The measurement procedure should be consistent with those for conventional treatment planning algorithms, for example, Fraass et al. (1998). Appropriate detectors should be used for the different situations, for example, measurements at shallow depths. Additionally, the detector type has to be carefully selected to receive a consistent measured data set. Possible dependencies in energy response, effective point of measurement, dose rate, and so on need to be taken into account. Furthermore, the voxel size for the calculated dose distributions should be approximately the same as the sensitive volume of the detector used for the measurements.

- Since MC calculations are always associated with statistical uncertainties, single-point dose comparisons are critical. For example, absolute dose calibration using a single point is inappropriate. A method using multiple points is much more robust (Siebers et al., 1999; Fix et al., 2007). This is also true for dose prescription in clinical treatment plans.

- Generally, it is advisable to perform dose comparisons at shallow depths, since these measurements are highly sensitive with respect to the beam model settings. Thus, they are useful to verify the accuracy of the beam model. Additionally, in-air profile measurements can be used for investigating the performance of the beam model since the impact of scatter is reduced.

- Although in the commissioning and validation ultimately the calculated dose distribution has to match measurements, it is important to bear in mind that the measurements themselves are also associated with errors.

Overall, the result of the commissioning and validation process strongly depends on the quality of the measurements. Figure 12.1 shows a typical comparison between MC calculated and measured absolute depth dose curves and lateral dose profiles. Using an incompatible voxel size for the calculation compared with the detector used for the measurement lead to dosimetric inaccuracies in the penumbra region. A study by Sahoo et al. (2008) shows that the penumbra width differs by a factor of 2 when different detectors are used. Figure 12.2 illustrates an example of the comparison between calculated and measured absolute dose profiles for 45° hard wedged beams. The comparison at different depths provides some indication whether or not the energy spectrum, including the beam hardening, has been characterized appropriately in the beam model. The validation of the MLC has to be done with great care since this complex beam modifier is often used in both static and dynamic treatment applications. Transmission, leakage, and various MLC-shaped fields should be included in the validation. Figure 12.3 demonstrates calculated and measured transmission and leakage profiles. If the interleave leakage is not taken into account for the MLC, the average transmission might be correct; however, the shape of the transmission profile cannot be reproduced. On the other hand, it is not clear if such a detailed modeling of the MLC is clinically needed. That dosimetric comparisons between dose calculation algorithms are important is demonstrated by a study by Reynaert et al. (2005). They encountered a 10% difference in dose volume histograms (DVHs) for the optical chiasm.
Monte Carlo Techniques in Radiation Therapy

when dose distributions are compared using Peregrine and their in-house MC system and assigned the obtained dose differences found to an inaccurate Elekta MLC modeling. Additional studies demonstrating the importance of accurate MLC modeling are described in the literature (Schwarz et al., 2003; Webster et al., 2007).

Besides static beam modifiers, there are also dynamic beam modifiers such as dynamic wedges or the MLC. For the EDW (dynamic wedge for Varian linear accelerators), one jaw of the secondary collimators is traveling from one side to the other while the beam is on, thus generating a wedged beam. Figure 12.4 shows some examples from the validation of EDWs for a 6 MV Varian beam. Measurements with an ionization chamber are taken at some points along the depth dose curves and dose profiles. Since the MLC can be used for both static and dynamic treatment applications, it is not enough to validate the static characteristics through transmission measurements as mentioned above. Furthermore, dynamic applications have to be validated. Figure 12.5 illustrates the comparison between MC calculated and measured dose values for one IMRT field from a verification plan using dynamic MLC, that is, one IMRT field is applied with gantry angle zero on a water phantom. For each measurement point using an ionization chamber (CC04), the whole IMRT field has to be delivered. Additionally, a corresponding film measurement has been performed in a homogeneous solid water phantom.

Apart from dose comparisons in homogeneous phantoms, it is also important to validate the treatment planning system in inhomogeneous phantoms. This allows the validation of the transport in nonwater materials (density and composition) as well as the characterization of the material through CT data sets. The first issue can be verified if digital phantoms are used for the calculation. Figure 12.6 shows an example of an inhomogeneous phantom comparing relative depth dose curves for a slab phantom with water–bone–water interfaces (Carrasco et al., 2007). A 6 and 18 MV beam with a 10 × 10 cm² field was used for the irradiation. Measurements are performed with TLDs, MOSFET, and several ionization chambers. These measurements are compared with calculations using different dose calculation algorithms, including MC, pencil beam, and convolution/superposition algorithms.

### 12.4 Research and Commercial MCTP Systems

#### 12.4.1 Research MCTP Systems

Over the years, many research institutions have implemented MC dose calculation algorithms. Those implementations have been coupled with different kinds of source models to allow photon beam MC treatment planning. These research planning systems are described in the literature and have been used to quantify the dosimetric difference between MC calculated dose distributions and those calculated with traditional treatment planning systems. Research planning systems developed at research institutions include the following:

- University of California, Los Angeles, RTMCNP (DeMarco et al., 1998): MC treatment planning based on MCNP4A using a speed-optimized photon particle transport and dose scoring within the standard lattice geometry. The RTMCNP preprocessor provides a user-friendly interface between the user and the MCNP4A command structure.
- Memorial Sloan-Kettering Cancer Center (Wang et al., 1998, 1999): An EGS4-based MC treatment planning environment using a dual-source beam model. The sources describe the primary and the scatter radiation of the treatment head, and are based on full MC simulation of the accelerator.

- Stanford University and Fox Chase Cancer Center, MCDOSE (Ma et al., 2002): This MC treatment planning system for electron and photon beams is coupled with the FOCUS system (Computerized Medical Systems (CMS), Inc., recently purchased by Elekta AB). MCDOSE is based

**FIGURE 12.5** (See color insert.) Comparison of MC calculated and measured dose distributions for one head and neck IMRT field using a 6 MV beam in units of cGy in water: (a) calculated dose distribution and measurement points (with dose values in cGy), (b) calculated dose profile and measurements along the white arrow, and (c) calculated dose profile and measurements along the black arrow.

**FIGURE 12.6** Comparison of measured and calculated relative depth dose curves in a water–bone–water phantom: 6 MV beam (left) and 18 MV beam (right). Measurements are: TLD, white circles; MOSFET, black squares; NACP02, black rhombus; and NE2571, white triangles. (Adapted from Carrasco P. et al. 2007. Med Phys 34: 3323–33, with permission from American Association of Physicists in Medicine.)
on EGS4 and as beam model either phase space files or multiple source models can be used.

- Virginia Commonwealth University (Siebers et al., 2000, 2002): An MC treatment planning system using EGSnrc as transport code. For the radiation transport through the MLC, a dedicated transport method has been developed (Siebers et al., 2002).
- University of Michigan, RT_DPM (Chetty et al., 2003): BEAMnrc phase space files coupled with the DPM MC code (Sempau et al., 2000) is used as dose calculation engine in this MC treatment planning system.
- University of Tübingen (Alber et al., 2003): An MC treatment system using MC also within the optimization. As beam model, the virtual fluence model (Fippel et al., 2003) is utilized and the XVMC code (Fippel, 1999) is used as a dose calculation algorithm. Recently, a dedicated source model for the electron contamination has been included in the beam model (Sikora and Alber, 2009).
- McGill University, MMCTP (Alexander et al., 2007): The McGill Monte Carlo treatment planning is a flexible software package with import options, for example, DICOM-RT, tools for contouring, beam editing, visualization, and analysis. MC dose calculations for electron and photon beams are supported using BEAMnrc for the simulation of the accelerator head and XVMC for the dose calculation.
- Inselspital and University of Bern, SMCP (Fix et al., 2007): The Swiss Monte Carlo Plan is an MC treatment planning system interfaced with Eclipse (Varian Medical Systems, Inc.). Transport methods of different complexity levels for the treatment head combined with MC dose calculation algorithms of EGSnrc or VMC++ allow automatic and flexible MC treatment planning.

12.4.2 Commercial MCTP Systems

In the recent decade, there have been only a few commercial treatment planning systems available offering MC dose calculation for photon beams:

- Peregrine is an MC dose calculation system developed at the Lawrence Livermore National Laboratory. Peregrine as implemented within the commercial treatment planning system Corvus (NOMOS Corporation) uses a four source beam model based on a phase space simulation of the accelerator head (Schach von Wittenau et al., 1999; Hartmann Siantar et al., 2001). Three photon subsources representing the target, the primary collimator, and the flattening filter are combined with the fourth source representing the electron contamination of the beam. A series of correlated histograms is used to reproduce the phase space information directly above the secondary collimator jaws. These sets of histograms are stored in a device file describing the accelerator according to the beam model and the beam modifiers (material, geometry, location, etc.). During the MC dose calculation, the particle is generated by sampling from the histograms and then transported through the beam modifier and the patient taking into account the backscatter into the monitor chamber. For the commissioning process, a whole class of possible beam models is provided by the Corvus system. Based on the linear accelerator specification provided by the user, a certain beam model object is created. The tuning of the beam model is performed by interpolation of the parameter data generated for different initial electron beam energies such that the calculated depth dose curve for a 10 \times 10 \text{ cm}^2 field matches the corresponding measurement. Additionally, the 40 \times 40 \text{ cm}^2 field is used for the commissioning. Validation of the Peregrine system for clinical use is shown by several groups (Hartmann Siantar et al., 2001; Heath et al., 2004; Lehmann et al., 2006; Rassiah-Szegedi et al., 2007).
- Monaco is the MC treatment planning option offered by CMS (now Elekta AB). The beam model in Monaco is based on the virtual fluence model for photons developed by Fippel et al. (2003) and for the electron contamination the approach developed by Sikora and Alber (2009) is used. The inputs for this beam model are depth dose curves, output factors, and lateral dose profiles at several depths. Commissioning uses small-field depth dose curves to determine the primary energy spectrum, output factors for a reference field as well as for the smallest and largest field size to determine the primary source diameter. Finally, cross profiles for the largest field size at several depths are needed to extract the off-axis energy fluence variations. This leads to a beam model that is characterized by 11 parameters. The radiation transport through the MLC is performed by using transmission filters in multiple layers (Sikora et al., 2007). This method results in a speed increase of about a factor of 100 compared with full MC transport and allows simulating the geometry of the MLC but not the scatter radiation produced in the MLC. For the patient dose calculation, the XVMC code is used (Fippel, 1999). Clinical validations have been reported in the literature (Sikora et al., 2007, 2009; Fotina et al., 2009; Sikora and Alber, 2009).
- In December 2008, Brainlab AG released their MC treatment planning system within iPlan. The beam model is based on a virtual beam model developed by Fippel et al. (2003) and supports linear accelerators, including their MLCs of all major vendors. For the MLC beam modifier, a full MC simulation is used as radiation transport. However, the user is able to select between a speed-optimized (simplified MLC without leakage) and an accuracy-optimized MLC model. Further parameters in iPlan are the voxel size for the dose calculation, the requested mean variance of the MC dose distribution, and whether dose to medium or dose to water is calculated. Currently, the MC dose calculation algorithm requires the commissioning of the pencil beam algorithm in iPlan. The commissioning and validation of the iPlan system is done by Brainlab itself using a total of 93 in-air measurements and 97 measurements in water. The in-air measurements include depth profiles, cross profiles, and output factors whereas the
measurements in water include absolute calibration measurements, depth dose curves, and cross profiles all at a source to surface distance (SSD) of 90 and 100 cm as well as output factor measurements at SSD 90. Most of these measurements in water are used for the validation of the beam model. The results of the commissioned beam model are finally provided to the customer. Recently, validations of the MC implementation in iPlan have been published in the literature (Fotina et al., 2009; Kunzler et al., 2009).

- In early 2011, the treatment planning system ISOgray (DOSIsoft) received FDA clearance. ISOgray offers MC dose calculations based on the general-purpose MC code PENELOPE (Salvat et al., 2006). The beam model consists of two parts: The first part includes the treatment head components of the linear accelerator that are patient independent beginning with the target and the second part is the patient-specific part, including the secondary collimator jaws and beam modifiers such as blocks, wedges, and MLC. For the patient-independent part, a phase space file is generated using PENELOPE. This computation is performed by DOSIsoft during the commission process in which the incident energy and the spot size of the initial electron beam impinging on the target are tuned to fulfill the acceptance criteria according to the measurements from the user. The phase space file is then used as input for the patient-dependent part of the treatment head where an efficient selective particle tracking method is used (Brualla et al., 2009). Thereby, the geometry of the beam modifier is divided into areas depending on whether or not secondary particles or scatter radiation can emerge from the beam modifier leading to skin and nonskin areas, respectively. Whereas in skin areas, the radiation transport is modeled accurately, in nonskin areas, the transport is simplified. In this manner, linear accelerator models for Elekta, Siemens, and Varian are available based on the according manufacturer information. For the dose calculation within the patient, PENFAST is used, which is based on PENELOPE but utilized particle radiation transport techniques that are optimized with respect to calculation speed (Habib et al., 2010). The work from Habib et al. (2010) also serves as a preclinical validation of this treatment planning system.

Furthermore, Nuclertron (now Elekta AB) has announced that MC photon treatment planning will be included in Oncentra in their upcoming releases.

12.5 Clinical Examples and Applications

12.5.1 MC as Treatment Planning Tool

There are many subjects that are worthwhile to be discussed in this section and most of them are described in detail in other chapters of this book. However, there are some issues that have to be carefully considered from a clinical perspective when using MC as a treatment planning tool. The two main issues involving the physicians’ point of view are discussed in the following sections.

12.5.1.1 Noise in Dose Distributions

While conventional dose calculation algorithms determine the 3D dose distributions deterministically, the MC method leads to a result associated with a statistical uncertainty. Thus, viewing and evaluating MC calculated dose distributions for the first time within a clinical environment is challenging and needs some time to get used to. The statistical uncertainty influences many quantities with respect to dose distributions such as isodose lines, DVHs, dose indices, convergence of cost functions, and so on. The statistical uncertainty is coupled with the number of independent particle tracks used for the dose calculation and is usually determined by using the history by history method (Walters et al., 2002). Roughly, the square number of histories is needed to halve the statistical uncertainty of the dose distribution. Figure 12.7 demonstrates the impact of the statistical uncertainty, that is, the number of independent histories, on the dose distribution for a lung case using three wedged beams. Thereby, the MC uncertainty is the average of the statistical uncertainties (added in quadrature) of all dose values in the dose distribution with more than 50% of its maximum dose. Although the difference in quality of the isodose line is significant, the differences are less pronounced in the DVHs as shown in the lower right plot for the three cases. Some differences for the planning target volume (PTV) and myelon are visible when the statistical uncertainty is reduced from 10% to 2%. However, in the DVH, there is virtually no difference if the statistical uncertainty is further reduced to 0.5%. From this example, it is clear that for MC dose calculations, all point values, such as $D_{\text{max}}$, $D_{\text{min}}$, and so on, are highly critical, since those values might be associated with a high statistical uncertainty in the considered calculation. The uncertainty might be different if the calculation is done a second time using different initial settings for the random number generator. This impact is also important for the dose prescription to a specific point. Thus, quantities relating to a volume or a larger number of dose values, such as $D_{\text{median}}$ or $D_{\text{mean}}$, are more reliable quantities when dealing with MC dose distributions. In general, a statistical uncertainty of 2% per beam results in a reasonable precision within the target volume, since usually three or more beams are used within a treatment plan. However, this might be no longer the case if the number of beams is below three or if OARs are considered. Since generally the OARs receive a lower particle fluence than the PTV, the statistical uncertainty in OARs can be much higher than in the target volume. Consequently, the evaluation for the OARs, for example, the DVHs or NTCP values, has to be done very carefully. It might be necessary to increase the number of histories to receive an acceptable statistical uncertainty in the normal tissue.

Other methods influencing the statistical noise are denoising techniques. Several different denoising algorithms have been developed. El Naqa et al. (2005) and the references therein provide a broad overview of existing denoising techniques as well
Monte Carlo Techniques in Radiation Therapy

12.5.1.2 Calculation Time

The statistical uncertainty and the number of histories needed for the simulation are directly related to the CPU time needed for the calculation of the dose distribution. However, there are also some additional factors affecting the dose calculation time, for example, the dose voxel size, the dose scoring volume, the used beam modifiers, and the CPU. For clinical applications, it is important to have fast dose calculation algorithms. A summary of timing results for clinical treatment plans are given in the Task Group Report 105 of the AAPM (Chetty et al., 2007). Bearing in mind that this timing comparison has been performed several years ago and that the calculation speed of the computers is still rapidly increasing, nowadays, the calculation time for photon MC treatment planning is acceptable for clinical routine. Apart from increased computer speed, the MC dose calculation codes themselves have been optimized using variance reduction methods (Chetty et al., 2007) or due to simplification of the transport code itself. For instance, the radiation transport within the beam modifier provides some potential in optimizing the calculation time, since there is also a trade-off between accuracy and CPU time. For example, the radiation transport through the secondary collimator jaws could be very simple (i.e., fast) if additionally the MLC is used below the jaws (Schmidhalter et al., 2010). Recently, MC algorithms have been implemented on graphical processing units demonstrating that the efficiency of the MC method could be further increased significantly (Badal and Badano, 2009; Jia et al., 2011). More information about this issue can be found in Chapter 19 of this book. Despite all these optimizations, the inverse planning might be the only exception where—given a large number of iterations—the overall calculation time might become unacceptable.

12.5.2 Comparisons of Dose Calculation Engines

12.5.2.1 Lung

Owing to the difficulties of the modeling of electron transport in low densities in the conventional dose calculation algorithms, lung is potentially the site showing the largest differences between dose distributions calculated using MC and conventional algorithms. There are many studies published in the literature investigating the differences between several
dose calculation algorithms. These include investigations using inhomogeneous phantoms (Carrasco et al., 2004; Fogliata et al., 2007, 2011; Panettieri et al., 2007; Aarup et al., 2009; Fotina et al., 2009; Bush et al., 2011) as well as patient treatment planning studies (Wang et al., 2002a; Dobler et al., 2006; Vanderstraeten et al., 2006; Hasenbalg et al., 2007; Madani et al., 2007). As an example for a phantom study, Fogliata et al. (2007) demonstrate that even for simple lung inhomogeneities, algorithms based on pencil beam convolution can lead to dose differences of up to 30% compared with MC. For more advanced algorithms, this difference is reduced to about 8%. This difference could even be further reduced for a grid-based Boltzmann equation solver (Bush et al., 2011; Fogliata et al., 2011). In summary, the studies with inhomogeneous phantoms show that the higher the energy and the smaller the field sizes, the larger the differences between MC and conventional dose calculation algorithms become. In terms of lung patient studies, Wang et al. (2002a) have shown differences of more than 10% when MC is compared with an algorithm using the equivalent path length (EPL) inhomogeneity correction method. Additionally, they report that due to the reduced lateral electron range for lower energies, using 6 MV photons instead of 15 MV is advantageous (Wang et al., 2002b). This has been confirmed in a more recent study by Madani et al. (2007). If highly accurate algorithms are available, the energy selection depends on priority ranking of endpoints. Vanderstraeten et al. (2006) compared MC calculated dose distributions for lung cancer patients with those calculated with the two different convolution/superposition algorithms Pinnacle (Philips Electronics N.V.) and Oncentra MasterPlan (Nucletron). Differences above 5% are reported. However, if the patient tissue is replaced by water, the differences above 5% disappear, thus demonstrating some limitations of dose calculation algorithms in conventional treatment planning algorithms. Hasenbalg et al. (2007) concluded that using a pencil beam algorithm may be inappropriate for lung cancer patients due to inaccurate calculated dose distributions. The more advanced dose algorithms AAA in Eclipse (Varian Medical Systems, Inc.) and collapsed cone convolution in MasterPlan (Nucletron) show dose differences of 5%. For stereotactic lung lesions, Dobler et al. (2006) show that the pencil beam algorithm overestimates measurements by up to 15% whereas the collapsed cone and MC underestimate measurements by 8% and 3%, respectively. In summary, the discrepancies concern differences in the build-up region, and the underestimation of the target coverage and of the isodose lines within the low-density tissue. The latter could also lead to differences in the dose distribution within the OARs themselves, which are located tangential to the incident fields. As an example with respect to the target coverage and isodose lines, Figure 12.8 shows the comparison between an MC and PB calculated dose distribution for a lung case within

![Figure 12.8](image-url)  
**FIGURE 12.8** Comparison of an MC and pencil beam (PB) calculated dose distribution for a lung case within iPlan: (a) MC isodoses, (b) PB isodoses, and (c) DVH comparison.
iPlan. The MC calculated dose distribution demonstrates that the dose coverage of the target is not achieved when using the PB algorithm for the dose calculation. This is also expressed in the DVHs shown in Figure 12.8 if for both dose calculation algorithms the same number of MUs is used. Whereas the dose to the PTV is overestimated when using the PB algorithm, the differences for the OARs are small.

12.5.2.2 Head and Neck

Although most comparison studies between MC and conventional dose calculation algorithms have been performed for lung, there is also a motivation of investigating this comparison for head and neck cancer patients. Since critical structures such as spinal cord are often located close to the target volumes, dose distributions with high dose gradients are applied. Additionally, air cavities are in these regions that lead to difficulties for many conventional dose calculation algorithms due to inaccurate dose calculations around air cavities, especially when Batho or ETAR correction methods are used (du Plessis et al., 2001). Examples of studies including investigations for head and neck cancer are the following: Wang et al. (2002a), Francescon et al. (2003), Seco et al. (2005), Boudreau et al. (2005), Knoos et al. (2006), Partridge et al. (2006), Sakthi et al. (2006), Mihaylov et al. (2007), and Zhao et al. (2008). The differences are smaller for head and neck compared with those for lung cancer and it has not been proven that these differences are clinically relevant. However, in cases with air cavities, differences may have to be carefully considered when choosing the dose calculation algorithm (Paelinck et al., 2006). As an example, Figure 12.9 shows a dynamic MLC IMRT larynx case. Comparisons between MC, AAA, and PB (Batho correction method applied) are shown together with the corresponding DVHs using the same number of MUs. The PB calculation results in a higher dose to the PTV compared with the AAA and MC algorithm. The MC method predicts a less homogeneous dose distribution within the PTV compared with the dose prediction when using the AAA algorithm. This observation corresponds with findings from other research groups (Seco et al., 2005; Sakthi et al., 2006). The dose to the OAR is predicted higher when MC is used. Overall, the differences are larger for the more simple pencil beam than for the more accurate AAA algorithm. Those differences need to be carefully considered in clinical routine.

12.5.3 Inclusion of Time Dependencies

One of the major advantages when using MC methods concerns the time dependencies of an irradiation in a 4D dose calculation. Whereas conventional algorithms usually discretize this problem, leading to increased calculation time, MC is able to sample the time with almost no increase in calculation time when considering the statistical uncertainty of the dose distribution in the target volume. Moreover, several approaches have been investigated to incorporate additionally the temporal patient movement into MC. One approach is to map the time-dependent dose distributions to a reference time point by using some kind of deformable registration between the time steps (Keall et al., 2004; Flampouri et al., 2006). Rosu et al. (2005) applied this method in conjunction with reducing the dose grid

FIGURE 12.9 (See color insert.) Comparison of an MC, AAA, and PB calculated dose distribution for a head and neck case: PB isodoses (a), AAA isodoses (b), MC isodoses (c), DVH comparison with PB (dots), AAA (triangles), and MC (squares) (d). The red lines show the PTV.
size to increase the accuracy of the method. Alternatively, the deformation could be taken into account by adjusting the voxel geometry, that is, using deformed voxel instead of a rectilinear dose grid (Heath et al., 2007; Peterhans et al., 2011). A method that separates the particle transport from the energy deposition scoring is the energy transfer method. Thereby, the particle transport is done in a rectilinear grid. The scoring, that is, the energy deposition, is done in a reference geometry at a certain time point using deformable image registration (Siebers and Zhong, 2008). A further development of this method is called the energy and mass congruent mapping (Zhong and Siebers, 2009). By this method, not only the energy deposition but also the mass is mapped to a reference geometry by deformable image registration. Apart from using a time to resolve the patient movements, there are other temporal dependencies that can be handled within the same simulation, for example, dynamic MLCs of an IMRT application and gantry rotation (Paganetti et al., 2004; Manser et al., 2010). This behavior opens the door not only to accurate dose calculation but also to a highly valuable tool for investigations such as interplay effects (Paganetti et al., 2005). Another example is reported in a study by Schmidhalter et al. (2007). They investigated the dosimetric effect in the case where the secondary collimator jaws follow the open window of the MLC during dynamic MLC IMRT to reduce transmission radiation through the MLC.

12.5.4 Reevaluation of Studies

In order to investigate the impact of more accurately calculated dose distributions on outcome, MC might be the most appropriate method for reevaluation of clinical studies. There are only a few studies reported in the literature dealing with lung cancer patients (Lindsay et al., 2007). For lung patients, the outcome analysis is difficult, since the treatment plans for many lung studies are lacking inhomogeneity corrections. However, as already mentioned, the impact of these corrections is large for low-density tissue; thus, it is very important to use accurate dose distributions for investigating the mapping to clinical outcome. These correlations between the dose distribution and effects for tumor and normal tissue gained from retrospective studies may also help to define protocols for prospective studies. Additionally, such retrospective studies provide the opportunity to learn how to use MC for routine treatment planning, since switching from one dose calculation algorithm to another is always a much debated issue in the clinical environment. This is mainly because the treatment protocols used are based on the clinical experience. Most likely the change of the dose calculation algorithm will also change the dose prescription, that is, a simple scaling, due to changes of the dose distribution in the tumor. However, the situation is more difficult for the normal tissue. Generally, the OARs are often located at the field edges and the impact of the dose calculation algorithm on their dose distribution due to different beam models, implemented lateral electron transport, and so on is not obvious, especially in low-density tissues. Consequently, accurate dose calculations in OARs are of critical importance, particularly if protocols for dose escalation are considered. Thus, many more studies are needed.

12.5.5 MC as QA Tool

Another application of MC dose calculations is its use as a QA tool. This could be done for different purposes. As a first example, MC dose calculations can be used as an independent MU calculation (Seco et al., 2005; Fix et al., 2007; Yamamoto et al., 2007; Pisaturo et al., 2009). Thereby, the treatment plan of the conventional algorithm is recalculated with MC and the resulting MUs are compared. Instead of using the MC as an independent MU check, it could be just used to recalculate some complex or unclear treatment plans, especially if it is assumed that for such a plan the conventional algorithm is likely to fail. Bearing in mind that these are some individual cases, not the calculation time but rather the accuracy is highly important. Electronic portal imaging systems (EPIDs) are also often used for IMRT QA procedures. Thereby, measurements performed with the EPID are compared with a reference. One approach is to use MC to determine the reference image for this comparison (Siebers et al., 2004; Frauchiger et al., 2007; Jarry and Verhaegen, 2007; Beck et al., 2009; Wang et al., 2009). Owing to the increased distance of the EPID to the photon source in the linear accelerator head and the high resolution of the EPID detector, the calculation time is very long and thus optimized transport methods have to be used, for example, MC calculated kernels (Wang et al., 2009). Another purpose for the use of MC dose calculation algorithms is benchmarking, since it is assumed that MC is the most accurate method to predict dose distributions (Vanderstraeten et al., 2006; Fix et al., 2007). Since for benchmarking, the accuracy is most important and there are generally only a few situations to be simulated, the calculation time is usually not very critical.

12.5.6 MC in Optimization

In the previous sections, the use of MC dose calculation for forward treatment planning has been described. However, in principle, the optimization also benefits from the accuracy achieved when using MC dose calculations in the inverse treatment planning process for IMRT or VMAT applications. In general, IMRT delivery is performed by a sequence of small fields or a dynamically moving and reshaping aperture. This leads to a great demand on the accuracy of the dose calculation algorithm. Furthermore, conventional dose calculation algorithms may give rise to dose errors, since a certain fraction of the dose is deposited by leakage and scatter radiation of the MLC. Most IMRT systems use simple but fast dose calculation algorithms, for example, pencil beam, within the optimization process. In many systems, a more sophisticated algorithm is used for the final dose calculation. The limited accuracy of these algorithms leads to dose prediction and convergence errors in the IMRT optimization (Mihaylov and Siebers, 2008).
Although the computing time of the MC dose calculation for a single field is very long compared with conventional dose calculation algorithms using a pencil beam or convolution/superposition, the MC calculation time is almost independent of the numbers of beams. This is not the case for the conventional dose calculation algorithms. However, the main disadvantage of using MC in the optimization remains to be the long calculation time, bearing in mind that many dose calculations during the optimization have to be done. As a consequence, one approach is to use MC dose calculation not in each iteration step but always after a certain number of iterations. In all other iteration steps, a fast dose calculation algorithm is used. This approach is called a hybrid scheme for IMRT optimization (Siebers et al., 2007; Siebers, 2008). Another approach is the use of MC calculated beamlets that are then used in the IMRT optimization process (Bergman et al., 2006; Sikora et al., 2009).

A few studies demonstrate that the use of MC in the optimization leads to improved dose distributions (Bergman et al., 2006; Dogan et al., 2006; Mihaylov and Siebers, 2008; Siebers, 2008; Sikora et al., 2009). Nevertheless, more detailed studies are needed to determine whether these improvements are clinically relevant.

### 12.6 Conclusions

MC dose calculation offers a great potential in photon treatment planning. Recently, the first few commercial treatment planning systems have MC methods for photon treatment planning available and more will come in the near future. Thus, MC dose calculation will be of increased importance in the future. Along with the shift to new dose calculation algorithms, there will be adjustments in clinical routine. It has been demonstrated that when using an MC dose calculation algorithm, not only the algorithm itself is changed, but almost the whole radiotherapy process is affected, for example, outlining, dose prescription, beam model, and CT conversion. There are still many open questions and studies required to carefully investigate all the impacts of this new technology and its impacts on outcome. This has to be done in close collaboration between clinical and research physicists.

### References


13.1 Introduction

Proton and ion beam therapies have gained importance recently with, at the time of this writing, 38 centers in operation and 22 under construction (http://ptcog.web.psi.ch/). Given the nature of curative radiotherapy (the aim of achieving as high a cure rate as possible with as low as possible severe side effects), uncertainty requirements for proton and ion beam dosimetry are as stringent as for any other form of radiotherapy. Proton and ion Monte Carlo simulations play a crucial role at different levels of the dosimetry chain: in primary standards and reference instruments for the calculation of perturbations and in relative dosimetry for the calculation of both perturbations and the energy dependence of many detectors.

With respect to electromagnetic interactions, the transport simulation of protons and ions is similar as for electrons with the simplifying condition that both scattering angles and energy straggling are much reduced. This makes multiple scattering and straggling theories developed for electron transport work well for protons and ions, and major Monte Carlo codes for the transport of protons and ions such as Geant4, MCNPX, and SHIELD-HIT use similar models. In general, artifacts induced by simplifications in the implementation of these theories, for example, in boundary crossing algorithms, are less severe as for electrons. Of course, if the transport of the delta electrons themselves plays an important role in a dosimetry problem, as for example in transport simulations through small cavities, accurate electron transport is still crucial and the same considerations as brought up in Chapter 5 hold. A major difference with electron transport is that nuclear interactions play an important role. These interactions can lead to break up of the target nuclei and in the case of ions also to break up of the projectile. The cross sections of these nuclear interactions are not well known (e.g., for protons, ICRU Report 63 (2000) quotes an uncertainty of 10% on the total interaction cross sections and 40% uncertainty on the production cross sections of secondary particles). Given that for a 200 MeV proton about 5% of the dose is deposited by charged secondary particles resulting from those nuclear interactions, the quoted uncertainties are potentially substantial for dosimetry problems.

This chapter reviews the application of proton and ion Monte Carlo simulations in the process of establishing primary standards, in calculating perturbations for ionization chambers in proton and carbon ion beam reference dosimetry, and in calculating absorbed dose response factors for relative dosimeters like alanine and radiochromic film using illustrations from the literature or new examples.

13.2 Primary Standards

Primary standards for proton and ion beam dosimetry can be divided in direct calorimetric methods and fluence-based methods (Palmans et al., 2009; Karger et al., 2010; Palmans, 2011).

In calorimetry, the temperature rise in the medium, $\Delta T_{\text{med}}$, as a result of the energy deposited by ionizing radiation is measured with high precision. Absorbed dose to the medium, $D_{\text{med}}$, is then obtained by multiplying with the specific heat capacity of the medium, $c_{\text{med}}$, and the necessary correction factors:

$$D_{\text{med}} = c_{\text{med}} \cdot \Delta T_{\text{med}} \cdot \frac{1}{1 - h} \cdot \Pi_k$$

(13.1)
where \( h \) is the heat defect that represents the (fractional) energy that has been deposited by ionizing radiation but is taken away from the medium by any changes of its physical or chemical state and \( \Pi_k \), is the product of correction factors for heat transported away or toward the measurement point, field nonuniformity, changed scatter, and attenuation due to the presence of nonmedium equivalent materials. The specific heat capacity is normally obtained from measuring the temperature rise in the medium due to electrical energy dissipation, either in a separate experiment or within the calorimeter setup. If the medium is water, the quantity of interest in reference dosimetry for radiotherapy, absorbed dose to water, is directly measured. If another medium is used, a conversion procedure is required.

Fluence-based methods rely on a measurement of the particle fluence at the surface, \( \Phi \), and knowledge of the stopping power in water at a shallow depth, \( z \):

\[
D_w(z) = \Phi \frac{S_w(z)}{P_w} \Pi_k, \tag{13.2}
\]

where \( S_w(z) \) is the stopping power of the proton spectrum at depth \( z \) and \( \Pi_k \) is the product of correction factors for beam divergence, scatter, field nonuniformity, beam contamination, and secondary particle buildup. It is obvious that this method relies on accurate values of the stopping power in water for which the uncertainty is estimated to be 1–2% for protons according to ICRU Report 49 (Palmons et al., 2009). The main instrument in use to measure the incident particle fluence is the Faraday cup, which enables an accurate measurement of the number of charged particles provided it is well designed to capture the entire beam and to repel charges generated in the entrance window and other components as well as to repel charges ejected from the cup back into it. For broad beams, an additional major uncertainty is due to the determination of the field area. For pencil beams, this uncertainty vanishes when the derived quantity is a laterally integrated dose. A major concern is the influence of electrons and protons generated in the entrance window that reach the collecting electrode (and thus contribute to the signal) as well as electrons liberated in and escaping from the collecting electrode (which enhance the signal). Both sources of perturbation to the measurement are usually suppressed by a guard electrode with a negative potential with respect to the electrode and casing sometimes with addition of a magnetic field.

### 13.2.1 Water Calorimetry

Water calorimeters require high-purity water to control the chemical heat defect (Ross and Klassen, 1996), which is usually contained in a small sealed glass or PMMA container. The perturbation by these container vessels can be measured or calculated by Monte Carlo simulations. The latter provides more information on the source of the perturbations. An example is shown here, where for three vessels used in proton dosimetry experiments (Palmons et al., 1996; Medin et al., 2006; Sarfehnia et al., 2010), these perturbations are calculated using MCNPX 2.4.0 (Pelowitz, 2005) and McPTRAN.CAVITY (Palmons, 2005), an in-house adaptation of PTRAN (Berger, 1993) that can simulate nonwater materials and inhomogeneous geometries. The vessels are a hollow PMMA cylinder of 40 mm diameter and 0.5 mm wall thickness (Palmons et al., 1996), a hollow glass cylinder of 67 mm diameter and 1 mm wall thickness (Medin et al., 2006), and a pancake cell of 24.9 mm thickness and 1.12 mm wall thickness (Sarfehnia et al., 2010). The resulting perturbation correction factors, obtained as the ratio of dose to water in a small volume around the measurement point in the absence and in the presence of the vessel, are shown in Figure 13.1 as a function of the residual range in water. For the PMMA vessels, the perturbations are negligible for all situations (note that PMMA is not preferred for other reasons (Seuntjens et al., 1993b)). For the glass vessels, the perturbation depends only on the thickness of the vessel, not on the shape, and a sigmoid fit to the data for the glass vessels is given by

\[
k_p = 0.555 + \frac{0.4455}{1 + e^{-3.5 [\log(R_{res}) - 0.75]}}, \tag{13.3}
\]

where \( R_{res} \) is the residual range.

The results are dominated by the shift of the effect depth of the measurement points due to the different water equivalent thickness of the glass vessel expressed as ratio of differential ranges in the continuous slowing down approximation (csda).

Sarfehnia et al. (2010) assumed this to be the only contribution and calculated this effect analytically. A small contribution, however, is also due to the different attenuation by nuclear interactions in the glass wall as compared to the equivalent water layer. This contribution was not more than 0.1% and the data in Figure 13.1 were corrected for it. Including this contribution, the simulated results are in good agreement with the measurement performed by Medin et al. (2006).

### 13.2.2 Graphite Calorimetry

#### 13.2.2.1 Volume and Gap Corrections

In graphite calorimetry, the diffusivity of the medium (graphite) is such that a sample (the core) of it has to be thermally isolated from the environment to be able to measure its temperature rise. This is usually achieved with vacuum gaps introducing two perturbations that are difficult to study experimentally: the gap effect and the volume averaging effect. Both were studied for a small-body graphite calorimeter in a 60-MeV low-energy proton beam (Palmons et al., 2004), where the effects are expected to be larger than in a high-energy proton beam, using variants of PTRAN that later led to the earlier-mentioned McPTRAN.CAVITY code. The correction factor for the gap effect can be defined in two ways. The most obvious is the substituted gap correction factor derived as the ratio of dose in the core with the gaps filled up with graphite and the dose in the core with the gaps present. A more favorable, but also more subtle, definition is the compensated gap correction factor, which is defined as the
ratio of the average dose over the core volume when the core is moved toward the inner surface of the gap facing toward the beam source and the (new) gap is filled up with graphite and the average dose when the gap is filled up with air. The method with the compensated gap assures that, apart from the difference due to the air gap, the attenuation the particles that enter the core have undergone is the same for both dose determinations in the ratio. The volume averaging correction factor, $k_{vol}$, accounting for the finite dimensions of the core, has been calculated as the ratio of the dose in a small sphere at the center of the core and the average dose over the core volume. The simulated geometries for the gap and volume averaging are shown in Figure 13.2.

The calculated correction factors are shown in Figure 13.3 for modulated and nonmodulated beams. For the modulated beams, both substituted and compensated gap corrections are smaller than 0.1% except when part of the core is in the distal fall-off region of the spread-out Bragg peak (SOBP). For the nonmodulated beam, the compensated gap correction is also small, but the substituted gap correction is substantial. The volume averaging corrections are smaller than a few tenths of a percent at shallow depth in a nonmodulated beam and in the plateau of the modulated beam. Nevertheless, the corrections and the variation of the latter might be larger than expected given the flat dose profile of an SOBP. This can be explained by ripples caused by the overlapping Bragg peaks illustrating that the dose at a point is very sensitive to small positioning errors.

### 13.2.2.2 Stopping Power Ratios

With a graphite calorimeter, one obtains dose to graphite from the measured radiation-induced temperature rise, while the quantity of interest is dose to water. A conversion procedure is thus needed. If the charged particle fluences, differential in energy at equivalent depths $z_g$ and $z_{w-eq}$ in a graphite phantom and a water phantom, respectively, are equal, then dose to graphite in the

![FIGURE 13.1](image1.png) Perturbation correction factors for glass and PMMA vessels used in past experiments calculated for monoenergetic proton beams with energies 50, 60, 100, 150, 200, and 250 MeV at 3 cm depth (hollow data points) and at depths halfway between the phantom surface and the Bragg peak (solid data points), calculated using MCNPX with type-A uncertainties of less than 0.2%. The thin curves represent data calculated for the same conditions using McPTRAN.CAVITY, which are almost indistinguishable from the MCNPX results. The thick smooth curve represents a sigmoid fit to all the data using glass vessels with a thickness of about 1 mm, regardless the shape.

![FIGURE 13.2](image2.png) Side view of the geometries for the Monte Carlo calculation of the substituted gap correction factor, compensated gap correction factor, and volume averaging correction factor. All geometries are axially symmetric around the beam axis; all material is graphite, except the shaded areas, which are either vacuum or graphite.
Monte Carlo Techniques in Radiation Therapy

graphite phantom, $D_g(z_g)$, and dose to water in the water phantom, $D_w(z_{w-eq})$, at the equivalent depth in water, are related by

$$D_w(z_{w-eq}) = D_g(z_g) \cdot s_{w,g}(\Phi_g),$$

(13.4)

where the water-to-graphite mass collision stopping power ratio for the total charged particle fluence distribution, $\Phi_g$, in graphite is

$$s_{w,g}(\Phi_g) = \frac{\sum_i \int_{E_{max}}^{E_{min}} \Phi_{E,g,i}(E) \cdot \left( \frac{S_{g,i}(E)}{\rho} \right) \cdot dE}{\sum_i \int_{E_{max}}^{E_{min}} \Phi_{E,g,i}(E) \cdot \left( \frac{S_{w,i}(E)}{\rho} \right) \cdot dE},$$

(13.5)

and $\Phi_{E,g,i}$ is the fluence distributions differential in energy for charged particle type $i$ in graphite and $S_{g,i}/\rho$ are the mass electronic collision stopping powers for particle type $i$. For a given depth in graphite, the equivalent depth in water is given by

$$z_{w-eq} = z_g \cdot \frac{r_{0,w}}{r_{0,g}},$$

(13.6)

where $r_{0,w}$ and $r_{0,g}$ are the proton ranges in graphite and water, respectively. Here $z_{80}$ is the depth on the distal edge of the Bragg peak where the ionization drops to 80% of the maximum ionization, was used as an estimate of the ranges $r_{0,w}$ and $r_{0,g}$.

Water-to-graphite mass collision stopping power ratios were calculated with Equation 13.5 using charged particle fluence distributions obtained from simulations using the Monte Carlo code Geant4 version 4.9.0 (Agostinelli et al., 2003) for 60 and 200 MeV monoenergetic protons incident on a slab phantom using the low-energy models for electromagnetic interactions, the default ICRU Report 49 stopping power parameterization and the precompound models for nonelastic nuclear interactions.
interactions. The stopping powers considered are the ones used in the Geant4 transport simulations, ICRU Report 49 stopping powers and stopping powers calculated using improved values of the mean excitation energies \( I = 82.5 \pm 2.5 \, \text{eV} \) for graphite from Burns (2009) and \( I = 80 \pm 2 \, \text{eV} \) for water from Paul (2006).

The resulting water-to-graphite stopping power ratios from the full Monte Carlo calculated fluence distributions are shown in Figure 13.4. The ICRU Report 49 data and the data based on the Burns and Paul \( I \) values agree within about 0.1% (but the uncertainties for the ICRU data are larger; about 0.9% vs. 0.5% for the Burns and Paul data) while Geant4 results deviate considerably. The Burns and Paul values are almost constant for all depths \( (s_{wg} = 1.116 \, \text{with an uncertainty of 0.5%}, \text{which can be regarded as the best estimate at present}) \).

13.2.2.3 Fluence Correction Factors

Since the fluence distributions in water and graphite at equivalent depths are not equal, a fluence correction factor \( k_f \) needs to be introduced in Equation 13.4:

\[
D_z(z_{w-c}) = D_z(z_g) \cdot s_{w-g} (\Phi_w) \cdot k_f, \tag{13.7}
\]

which, given the relation between dose and stopping power in both materials, can be calculated as

\[
k_f = \sum_i \int_{0}^{E_{\text{max}}} \Phi_{E, i} (E) \cdot \frac{S_z (E)}{\rho} \cdot dE
\]

In these expressions, \( \Phi_{E, i} \) and \( S_z \) are the fluence distributions differential in energy for charged particle type \( i \) in water or graphite, respectively. The fluence correction factor defined in Equation 13.8 can be interpreted as the conversion of dose to water in the graphite phantom, to dose to water at an equivalent depth in the water phantom.

Simulations were performed using Geant4 version 4.9.0 (Agostinelli et al., 2003) using the same geometry and settings as for the stopping power simulations. The calculated fluence correction factors are shown in Figure 13.5. For both energies, there is a small difference between the simulation in which only protons are considered or the one in which all secondary charged particles from nuclear interactions are considered (the difference being dominated by the contribution from alpha particles). The general trends are that the fluence correction factor is below unity at shallow depths and that it increases over the entire range of the beam. For 60-MeV protons, the value for water equivalent depths up to 2 g cm\(^{-2}\) is consistent with an experimental value of \( k_f = 1.000 \pm 0.003 \) determined in a narrow beam setup for the same energy (Palmans et al., 2011). For 200 MeV protons, the fluence correction factor is varying more substantially and is about 2–3% below unity for depths up to about 10 g cm\(^{-2}\) beyond which it increases steeply to 6% above unity at a depth of 25 g cm\(^{-2}\).

13.2.3 Faraday Cup Dosimetry

One of the difficulties with Faraday cups is the suppression of signal currents from secondary electrons generated in the cup that manage to escape as well as from secondary electrons and protons generated in the entrance window. While these are all as much as possible suppressed by electromagnetic fields, Monte Carlo simulations can help in understanding and quantifying these corrections as well as feeding into design considerations.
Monte Carlo Techniques in Radiation Therapy

for Faraday cups. Geant4 (Agostinelli et al., 2003) has the capability of combining transport of all particles with accelerations and deflections of charged particles by electromagnetic fields.

Another application of interest is the use of multilayer Faraday cups (MLFC) to quantify the number of nuclear interactions taking place in slabs of nonconducting materials by using the principle of mirror charges. The charge collected in each of a series of thin metallic foils separated by the insulating absorber plates is measured and the method basically determines in which slab protons stop (by nuclear interactions in the entrance plateau and at the end of their range due to electromagnetic interactions in the Bragg peak). This can serve as a test for both nuclear interaction and energy straggling models in the Monte Carlo simulation of the device and the experimental data from such an experiment (Paganetti and Gottschalk, 2003) have been proposed as a benchmark with this purpose (Palmans and Capote-Noy, 2010). Figure 13.6 gives the comparison of the experimental results with Geant4 simulations (Paganetti and Gottschalk, 2003), FLUKA (Ferrari et al., 2005; Battistoni et al., 2007) simulations (Rinaldi et al., 2011), and SHIELD-HIT (Sobolevsky, 2011) simulations (Henkner et al., 2009b). The agreement in the Bragg peak is excellent in all cases but deficiencies in the nuclear models in different codes (or different nuclear interaction models in the same code, Geant4) are visible in the entrance regions. The experimental data points beyond the Bragg peak could have a contribution from measurement noise as well as from knock-on protons generated by neutrons. The substantial disagreement between the measured data and all models cast doubt on the idea that the latter contribution would be significant.

13.3 Ionization Chambers

Reference dosimetry using ionization chambers is recommended to be performed according to IAEA TRS-398 (2000).

Absorbed dose to water, $D_{w,Q}$, in a proton or ion beam with quality $Q$ is derived as

$$D_{w,Q} = M_Q \cdot N_{w,Q} \cdot k_{Q,0b}$$  \hspace{1cm} (13.9)

where $M_Q$ is the ionization chamber reading corrected for influence quantities, $N_{w,Q}$ is the absorbed dose to water calibration coefficient of the ionization chamber in a calibration beam of quality $Q_0$, and $k_{Q,0b}$ is the beam quality correction factor accounting for the use of the calibration coefficient in a different beam quality $Q$.

Owing to the limited availability of experimental data, $k_{Q,0b}$ values are in practice calculated from

$$k_{Q,0b} = \frac{(W_{air}/\varepsilon_Q) \cdot (s_{w,air})_Q \cdot p_Q}{(W_{air}/\varepsilon_{Q_0})_Q \cdot (s_{w,air})_{Q_0} \cdot p_{Q_0}}$$  \hspace{1cm} (13.10)

where, for both beam qualities, $W_{air}$ is the mean energies required to produce an ion pair in dry air, $s_{w,air}$ is the water-to-air mass stopping power ratios for the local charged particle spectrum, and $p_Q$ and $p_{Q_0}$ are the ionization chamber perturbation correction factors in both proton/ion and calibration beam. The beam quality index $Q$ for protons is the residual range $R_{res}$ defined as

$$R_{res} = R_p - z$$  \hspace{1cm} (13.11)

where $R_p$ is the practical range, defined as the depth distal to the Bragg peak at which the dose is reduced to 10% of its maximum value (for ions with a significant fragmentation tail, an estimate for the primary particle peak is made by constructing a tangent to the distal Bragg peak fall-off), and $z$ is the depth of measurement. $R_{res}$ is related to the most probable energy of the highest proton energy peak in the spectrum.
Monte Carlo Calculations for Proton and Ion Beam Dosimetry

While the product of $W_{\text{air}}$ and $s_{w,\text{air}}$ is mainly derived from experimental comparisons of calorimeters and ionization chambers, Monte Carlo simulations have mainly aided in calculating accurate values of the water-to-air stopping power ratios and perturbation correction factors, which is the subject of the following subsections.

### 13.3.1 Stopping Power Ratios

Monte Carlo simulation offers the advantage of in-line calculation of stopping powers and cavity integrals. This was illustrated for the dose conversion between graphite and water before. Medin and Andreo (1997) used this method to calculate water-to-air mass collision stopping power ratios for ionization chamber dosimetry in clinical proton beams. They used the in-house code PETRA, which includes transport of secondary electrons and secondary protons. Figure 13.7 gives the curves as a function of the difference between the csda range and the measurement depth, which is close to the residual range. The Spencer–Attix stopping power ratios are found to be about 0.5% higher than unrestricted stopping power ratios for the same proton spectra in high-energy protons. Secondary protons have a minimal influence. The variation of the stopping power ratio as a function of energy is not more than 0.5% from $R_{\text{res}} = 0.5$ cm to the highest values encountered in clinical proton beams.

SHIELD-HIT (Sobolevsky, 2011) has been used to calculate water-to-air stopping power ratios for ionization chamber dosimetry in clinical carbon ion beams (and for other ions) (Henkner et al., 2009a; Lühr et al., 2011). The results as a function of residual range are shown in Figure 13.8. Both studies for proton and carbon ion beams show that the residual range is a good beam quality specifier corresponding with unique values of the stopping power ratio for depths shallower than the practical range $R_p$.

![Figure 13.6](image_url)  
**FIGURE 13.6** Comparison of the experimental data from the multilayer Faraday cup (Paganetti and Gottschalk, 2003) (symbols) with the simulations from three different Monte Carlo codes (full black line: Geant4 (Paganetti and Gottschalk, 2003), dotted black line: FLUKA (Rinaldi et al., 2011), full gray line: SHIELD-HIT (Henkner et al., 2009b)). (a) Shows the charge per $10^9$ protons collected in each metallic plate, (b) shows the ratios between the Monte Carlo calculated data and the experimental data.
13.3.2 Primary Particle Fluence Perturbation

The change of the proton fluence from a point in water to the cavity volume of an ionization chamber can be modeled as a gradient correction or alternatively as a shift of the effective depth in water from the center of the chamber. For carbon ions, projectile fragments need to be accounted for as well, but it can be assumed that the local changes of these fluence spectra over the dimensions of an ionization chamber are small. A detailed analytical model tracking particles along straight lines through the ion chamber geometry and averaging the depth corresponding with the kinetic energy they have in the cavity was used to calculate the effective depth of cylindrical ionization chambers in proton and carbon ion beams (Palmans, 2006). Extensive Monte Carlo simulations using McPTRAN.CAVITY confirmed the analytical model. These were performed by simulating dose in the air cavity as a function of depth for all cylindrical ionization chamber types tabulated in IAEA TRS-398 for proton energies from 60 to 200 MeV and compare this with dose to air at the reference points in a homogeneous water phantom. All ionization chambers and all depths were simulated in one run on a distributed computing...
Monte Carlo Calculations for Proton and Ion Beam Dosimetry

grid system by implementing geometry interrogation, history splitting, and lateral range rejection. The geometry of the calculation is shown in Figure 13.9.

Figure 13.10 shows how the shift of the ionization chamber’s effective depth from the chamber center, \((-\Delta z_{\text{w, tot}})\), can be derived from the Monte Carlo simulations by calculating the distance from the center of the ionization chamber to the depth where dose to air in homogeneous water (in a point) equals the average dose to air in the air cavity. This can be done on the proximal slope (plateau) of the depth dose curve provided the curvature is sufficiently small or on the distal edge beyond the Bragg peak region.

Table 13.1 shows the values of \((-\Delta z_{\text{w, tot}})\) obtained from the analytical model and from the Monte Carlo simulations using both methods explained in the two previous paragraphs. The selection of ionization chamber types listed in Table 13.1 have all been reported in the literature as being used as reference instruments. The results are compared with the general recommendation in IAEA TRS-398 to use 0.75 times the radius of the cavity. The last column shows experimental determinations of the effective depth in water showing that in all cases, the agreement is within the experimental uncertainties.

Both the frontal and distal results are in general in good agreement with the analytical results. The deviation from the IAEA recommendation is small for most ionization chambers but can be considerable (up to 1.5 mm) for some chambers commonly used in proton dosimetry (Palmans, 2006.) Among the chambers listed in Table 13.1, the largest deviation is for the FWT IC-18. For Farmer-type ionization chambers as well as for most ionization chambers with a thin wall and a central electrode diameter that is small compared to the cavity diameter, the agreement is in general good (Jakel et al., 2000; Palmans et al., 2000).

The possibility of applying the analytical model to the more challenging problem of converting a depth dose curve obtained with an ionization chamber to the depth dose in homogeneous water is demonstrated in Figure 13.11. A function consisting of an arbitrary set of depth dose values connected by a cubic spline was constructed. This function was convolved with the analytical model and a least square optimization of the set of depth dose values was performed to minimize the difference with the Monte Carlo simulated depth dose function for the ionization chamber. This method is illustrated for the NE2581 in Figure 13.11, which shows the optimized data points and the difference of the cubic spline with the depth dose curve in homogeneous water. This results in differences to the real depth dose curve smaller than 1%.

of the dose maximum in the region frontal to the Bragg peak and 5% in the distal edge, where these differences are less important since they correspond to minor shifts in depth. This indicates that the method will be applicable to measured depth dose curves, which will always exhibit a certain level of fluctuations, as well.

13.3.3 Secondary Electron Perturbation

Several Monte Carlo studies using EGS4 to track secondary electrons in the past have demonstrated that secondary electron perturbations for ionization chambers in proton beams are not negligible and can amount to 1% and more for some wall materials (Palmans and Verhaegen, 1998; Palmans et al., 2001; Verhaegen and Palmans, 2001). In a more recent study, similar calculations were performed given that now better electron transport is implemented in EGSnrc and more computing resources are available enabling the calculation of these effects with lower uncertainties (Palmans, 2011). These are shown here as an example.

The modeled geometry is an air cylinder with a length of 24 mm and a diameter of 6.4 mm surrounded by a wall with a thickness of 0.38 mm, that is, dimensions close to that of a Farmer-type chamber, immersed in water. Both graphite and A150 as wall material were modeled. The broad monoenergetic

![Figure 13.11](image)

**FIGURE 13.11** Reconstruction of a depth dose curve (circles) derived from a Monte Carlo simulated distribution in depth of dose to air in the cavity of a NE2581 ionization chamber (+ symbols) with deliberately added noise (1σ = 2%) in a 60 MeV monoenergetic beam. The full curve is the Monte Carlo calculated depth dose curve in homogeneous water. (Adapted from Palmans H. Perturbation factors for cylindrical ionization chambers in proton beams. Part I: Corrections for gradients. Phys Med Biol 2006; 51(14): 3483–3501. With permission.)
Monte Carlo Calculations for Proton and Ion Beam Dosimetry

A proton beam was incident perpendicular to the chamber axis without any loss of energy over the geometry. Electron recoil spectra in the cavity, wall, and surrounding water were calculated from the Rutherford cross section and the number of interaction centers derived from the electron density in each medium. Isotropic emission of electrons was considered, which is justified given the rotational symmetry of the problem. Electron transport through the geometry was simulated using DOSRZnrc (Rogers et al., 2010) with default transport and boundary crossing settings and cutoff energy for electron transport of 1 keV.

Dose to air in the cavity, $D_{\text{air,cav}}$, for a broad proton field was evaluated as

$$
D_{\text{air,cav}} = \Phi_p \left( \frac{S^A}{P^P} \right)_{\text{air}} + \Phi_p N_A \left( \frac{Z}{A} \right)_{\text{air}} \times \left[ \int_{E_{\text{dep,MC}}} W_{\text{air}} \cdot \sigma_W \cdot dW \right] \cdot E_{\text{dep,MC}} \quad (13.12)
$$

where $\Phi_p$ is the (monoenergetic) proton fluence, $(S^A/P^P)_{\text{air}}$ the restricted mass stopping power at the energy of the protons, $N_A$ the Avogadro’s number, $(Z/A)_{\text{air}}$ the average ratio of atomic number and atomic weight of the medium, $W_{\text{max}}$ the maximum energy that can be transferred to a secondary electron by the protons, $\sigma_W$ the Rutherford cross-section differential in energy transfer (the integral thus being the total cross section for emission of an electron), and $E_{\text{dep,MC}}$ the Monte Carlo calculated dose contributions by electrons per electron set in motion in the cavity.

Wall perturbation correction factors were obtained as the ratio of $D_{\text{air,cav}}$ calculated by expression (13.12) for the chamber geometry with the wall set to water and for the chamber geometry with the wall set to the actual wall material. The calculated wall perturbation correction factors as a function of energy are shown in Figure 13.12. For both wall materials, the correction factors exponentially saturate at high energies, for A150 around 1.0045 and for a graphite wall around 0.985. The ratio for the A150 and graphite-walled chamber at 55 MeV is 1.0045, which is in close agreement with the reported experimental value of $1.005 \pm 0.002$ in a nonmodulated beam of the same effective energy (obtained by comparing a NE2571 with an in-house built A-150 chamber with the same dimensions) (Palmans et al., 2001).

### 13.4 Energy Dependence of Relative Dosimeters

Many solid-state dosimeters exhibit an energy-dependent dose response for protons and ions. If this energy dependence is known, it is sufficient to know the local charged particle fluence differential in energy to predict the change in response of the dosimeter in the phantom or to correct its response for this energy dependence.

The relative effectiveness $RE_i$ for charged particle type $i$ can be defined in terms of an isoresponse dose as (Bassler et al., 2008; Herrmann et al., 2011)

$$
RE_i = \frac{D_{\text{det}}}{D_{\text{air,i}}} \quad (13.13)
$$

where $D_{\text{det}}$ is the dose given to the detector by charged particle type $i$, and $D_{\text{air,i}}$ is the dose that needs to be delivered in a $^{60}$Co calibration beam to induce the same detector reading as the dose $D_{\text{det}}$. $RE_i$ is energy dependent.

Assuming that only charged particles contribute to the local dose deposition, the average relative effectiveness for the local spectrum can be derived from the charged particle fluence differential in energy $\Phi_p$ for all particle types $i$ as
\[
\frac{\sum \int_{0}^{\infty} RE_i(E) \cdot \Phi_{E_i} \cdot \left( \frac{S}{\rho} \right)_{\text{det},i} \cdot dE}{\sum \int_{0}^{\infty} \Phi_{E_i} \cdot \left( \frac{S}{\rho} \right)_{\text{det},i} \cdot dE}
\]

where \((S/\rho)_{\text{det},i}\) is the mass stopping power of the detector material for particle type \(i\). Monte Carlo is the ideal tool to calculate these fluence distributions. Two examples are given here for radiochromic film and alanine.

### 13.4.1 Radiochromic Film

The energy dependence of two types of radiochromic film (GafChromic MD-V2-55 and EBT) was measured by Kirby et al. (2010) by comparison of their response with that of a calibrated Markus ionization chamber. Depth dose curves of 15 and 29 MeV proton beams calculated with FLUKA and the predicted underresponse according to Equation 13.14 are shown in Figure 13.13 and compared with the experimental depth response data recorded with MD-V2-55 film. The films were stacked perpendicular to the beam direction and the data points correspond with the depth of the sensitive layer of the films. The energy dependence of the film response was used in a study of laser-induced protons to infer the energy spectrum from the measured depth dose curve (Kirby et al., 2011). A library of depth dose curves and corresponding fluence spectra differential in energy as a function of depth was calculated using FLUKA and the spectrum was deconvolved from the measured depth response curve. Accounting for the reduced relative effectiveness of the film, the proton fluence inferred from this method was found to be 30% higher than when not taking it into account, as was the practice used in previous papers, demonstrating that it is essential to make those corrections for the energy dependence of the dose response.

### 13.4.2 Alanine

The relative effectiveness of alanine in protons and ions was obtained from literature data (Palmans, 2003; Bassler et al., 2008). This information was used to predict the underresponse of alanine in the measurement of depth dose distributions in carbon ion beams and compare this with experimental data (Herrmann et al., 2011). The simulations were performed with FLUKA implementing the exact geometry of the stack of alanine pellets with a diameter of 5 mm and nominal thicknesses of 2.5 mm. The predicted response of alanine was derived from the fluence distributions differential in energy according to Equation 13.12. The comparison, exhibiting excellent agreement, is shown in Figure 13.14 for an SOBP established by the use of a ripple filter, which was modeled in detail in the simulation as well (Herrmann et al., 2011). As for radiochromic films, this work shows that it is essential to correct for the relative effectiveness when measuring depth dose curves using alanine.

### 13.5 Summary and Outlook

Monte Carlo simulations are an important tool in all levels of the dosimetry chain for proton and ion beam radiotherapy.

In primary standard calorimeters that aim to measure dose at a point in a homogeneous medium, Monte Carlo simulations are used to determine perturbations for any deviations from those conditions such as the presence of the high-purity-water-containing glass vessel in water calorimeters and the presence of vacuum gaps in graphite calorimeters as well as the fact that dose

![FIGURE 13.13 Depth dose curves for 15 and 29 MeV protons in PMMA simulated with FLUKA (solid lines) and corresponding depth dose reading (or isoeffect dose) curves for GafChromic MD-V2-55 (dashed lines) compared with experimental dose response data (squares). (Reproduced from Kirby D et al. Radiochromic film spectroscopy of laser-accelerated proton beams using the FLUKA code and dosimetry traceable to primary standards. Laser Part. Beams 2011; 29:231–239. With permission.)](image-url)
is measured over a volume in the latter. In graphite calorimetry, the calculation of stopping power ratios and fluence corrections needed for an accurate conversion between dose to graphite and dose to water also rely on Monte Carlo simulations.

For clinical reference dosimetry using ionization chambers, Monte Carlo simulations have made a major contribution to the calculation of water-to-air stopping power ratios and in the calculation of factors to correct for perturbations due to the presence of gradients and due to secondary electrons generated in the wall and the central electrode.

Since most solid-state detectors exhibit an energy-dependent dose response, it is essential to model the influence of the local spectrum when using them for measuring dose distributions. Examples are given of the application of Monte Carlo to this issue for alanine and radiochromic film and similar approaches could be used for other solid-state detectors.

Monte Carlo simulations will continue to play an essential role in all these aspects of dosimetry. The increased use of scanned particle beams will require characterizing calorimeters, ionization chambers, and relative dosimeters for their dose response in these complex dose delivery schemes. The Monte Carlo method is the only approach that can accurately model this complexity. The anticipated increased usage of solid-state detectors for treatment planning verification and auditing purposes in proton and ion beam dosimetry require corrections for the energy dependence of the dose response and thus knowledge of local charged particle spectra for which Monte Carlo is ideally suited. Also, the fluence correction factors potentially required when using plastic phantoms can be assessed by Monte Carlo in the same ways as for the conversion of dose to graphite derived from graphite calorimetry to dose to water.

Nuclear interactions are of importance in various dosimetric problems, such as the fluence corrections discussed above. In most cases, however, their influence is rather modest. An exception to this is PET and prompt gamma imaging (treated in Chapter 4) where the relations between dose distributions and the measured signals are critically dependent on accurate nuclear interaction cross sections, especially for the reaction channels producing the radioactive isotopes. Monte Carlo simulations will play an essential role not only in deriving dose distributions from the measurements but also in the benchmarking of the nuclear interaction models. An IAEA report reviewing the importance of nuclear interactions in dosimetry and providing such benchmark exercises will be forthcoming (Palmans and Capote-Noy, 2010).

References
Burns DT. A re-evaluation of the I-value for graphite based on an analysis of recent work on W, S(c,a) and cavity perturbation corrections. Metrologia 2009; 46: 585–590.


http://ptcog.web.psi.ch/.


14.1 Introduction

14.1.1 Short Introduction to Proton Therapy Beam Delivery

Most proton therapy facilities are currently based on the passive scattering technique. In the future, it is expected that more and more proton therapy patients will be treated using magnetic beam scanning. Both techniques can deliver a spread-out Bragg peak (SOBP) with a homogeneous target dose for each field, while the latter technique also allows intensity-modulated proton beam therapy. The basic principles are outlined in Figure 14.1 (see Paganetti and Bortfeld, 2005 for details).

To deliver a uniform dose to the target for a given treatment field (i.e., beam angle), an SOBP ensures coverage along the thickness of the target volume in the patient. It is created by combining pristine Bragg curves with different beam energies entering the patient. This can be achieved using a rotating absorber consisting of steps of various water-equivalent thicknesses (as in passive scattering) or by discretely changing the beam energy for each energy layer (as in beam scanning). In some scanning systems, the beam energy is adjusted using a range shifter in the treatment head. To cover the target volume laterally, the beam needs to be scanned magnetically in the x and y directions (beam scanning) or the proton beam needs to be broadened by using scatterers and absorbers consisting of various material combinations to achieve a laterally flat dose profile (passive scattering). In passive scattering, further shaping of the field is accomplished by apertures that confine the lateral dimension of the field and range compensators that are sculpted to conform the dose distribution to the distal shape of the target volume. For beam scanning, no patient-specific hardware is needed. However, depending on the width of the delivered pencil beams, an aperture might be used to improve the lateral penumbra of the beam.

14.1.2 Proton Physics

The energy loss of protons can be calculated using the Bethe-Bloch equation. Monte Carlo simulations implement multiple scattering in condensed history class II methods (Kawrakow and Bielajew, 1998). The Moliere theory predicts the scattering angle distribution and the Lewis method (Lewis, 1950) allows calculation of moments of lateral displacement, angular deflection, and correlations of these quantities. The implementation of the multiple scattering theory can differ slightly from Moliere's theory (Andreo et al., 1993; Gottschalk et al., 1993; Urban, 2002). Multiple coulomb scattering is the main cause for a broadening of the beam with depth.
The majority of dose is deposited via ionization, excitation, multiple coulomb scattering, and nuclear interactions. Depending on the interaction type, the physics is based on models, parameterizations, experimental data, as well as combinations of these. For nuclear interactions, cross sections as a function of proton energy may not be available for all reaction channels and need to be approximated. Although nuclear interactions do not significantly influence the shape of the Bragg peak, they do have a significant impact on the depth dose distribution because they cause a reduction in the proton fluence as a function of depth as well as secondary protons contributing significantly mainly in the entrance region of the Bragg curve (Paganetti, 2002). Note that as a rule of thumb, about 1% of the primary protons undergo a nuclear interaction per centimeter beam range. In pencil beam scanning, secondary particles emitted in nuclear interactions cause a “nuclear halo” around each pencil. The contribution can be significant when adding multiple pencils (Soukup and Alber, 2007; Sawakuchi et al., 2010). Interactions between the proton and a nucleus can be modeled as an intranuclear cascade with the probability of secondary particle emission. Once the energy of the particles in a cascade has reached a lower limit, a preequilibrium model can be applied.

Production cuts for secondary particles can influence energy loss and thus the simulation results (van Goethem et al., 2009). For proton dose calculation, primary and secondary protons account for about 98% of the dose, depending on the beam energy and including the energy lost via secondary electrons created by ionizations (Paganetti, 2002). The maximum range of the delta electrons in water is about 2.5 mm for a 250 MeV proton. The highest energy electrons are preferentially ejected in a forward direction. The energy of most electrons in a proton beam is much less than 300 keV, which corresponds to a range of 1 mm in water. The explicit tracking of secondary electrons can thus be neglected in most cases. For applications other than dose calculation, additional particles might need to be considered, for example, neutrons.

A Monte Carlo code might allow different physics settings from which the user can choose. Some settings might not be tailored for the energy domain of proton therapy and therefore require adjustment (Herault et al., 2005; Stankovskiy et al., 2009; Pia et al., 2010). The dependencies of proton therapy-related simulations on different physics settings can be considerable (Paganetti and Gottschalk, 2003; Kimstrand et al., 2008a; Zacharatou Jarlskog and Paganetti, 2008a). Even with the
“Correct” settings, values for the mean excitation energy might have uncertainties in the order of 5–15%. Such an uncertainty for beam-shaping materials can lead to up to 1 mm uncertainty in the predicted beam range in water (Andreou, 2009). It is important to consider this uncertainty when simulating energy loss in thick targets (Moyers et al., 2007; Andreou, 2009; van Goethem et al., 2009; Gottschalk, 2010). In particular for nuclear interactions, there are significant uncertainties in physics cross sections or models. For example, according to the international commission on radiation units & measurements (ICRU) (ICRU, 2000), angle-integrated emission spectra for neutron and proton interactions are known only to within 20–30% uncertainty.

### 14.1.3 Proton Monte Carlo Systems

The Monte Carlo codes widely used in proton therapy are FLUKA (Ferrari et al., 2005; Battistoni et al., 2006), Geant4 (Moyers et al., 2007; Gottschalk, 2010), MCNPX (Waters, 2002; Pelowitz, 2005), VMCP (Fippel and Soukup, 2004), and Shield-Hit (Dementiev and Sobolevsky, 1999). These codes differ not only in accuracy (presumably this difference is small) but also in the user interface. For example, Geant4 provides only an assembly of object-oriented toolkit libraries with the functionality to simulate different process organized in different functions within a C++ class structure. The codes also differ in the level of control over tracking parameters (e.g., physics settings, step sizes, and material constants).

In addition to the codes mentioned above, there are also programs that serve as interfaces to Monte Carlo codes. For Geant4, these are, for example, PTsim (Aso et al., 2010), GATE (Santin et al., 2003; Jan et al., 2004), GAMOS (Arce et al., 2008), and TOPAs (Perl et al., 2012). TOPAs (TTool for PArticle Simulations), makes Monte Carlo simulation more readily available for research and clinical physicists aiming at modeling passive scattering or scanning beam treatment heads. Furthermore, the user is able to model a patient geometry based on computed tomography (CT). The code provides advanced graphics, and is fully four-dimensional (4D) to handle variations in beam delivery and patient geometry during treatment. All this is achieved with a user-friendly interface that does not require experience in programming.

### 14.2 Simulating a Radiation Field Incident on a Patient or an Experimental Setup

#### 14.2.1 Characterizing Proton Beams at the Treatment Head Entrance

Proton therapy facilities typically consist of a cyclotron or synchrotron, which is connected to one or several treatment rooms by a complex beamline. The basic beamline elements between the accelerator and the treatment room that might be simulated using Monte Carlo are bending magnets and energy degraders. Monte Carlo beam transport through carbon and beryllium degraders has been performed with the goal of improving beam characteristics (van Goethem et al., 2009). However, beam optics calculations are often done numerically (Kurihara et al., 1983) despite specialized Monte Carlo codes that simulate magnetic beam steering (Brown et al., 1980).

Thus, Monte Carlo simulations typically start either at the exit of the treatment head, if a phase space parameterization or beam model is defined, or at the entrance of the beamline into the treatment room. If the Monte Carlo simulation starts at the treatment room entrance, a reliable parameterization of the beam at this position is needed. The variables are beam energy \(E\), energy spread \(\Delta E\), beam spot size \(\sigma_x, \sigma_y\), and beam angular distribution \(\sigma_{\theta_x}, \sigma_{\theta_y}\) (Paganetti et al., 2004a; Hsi et al., 2009). There are various methods to measure these parameters. The beam spot size can usually be determined with a segmented transmission ionization chamber located at the treatment head entrance. The size of a proton beam is usually in the order of 2–8 mm in \(\sigma_x\) or \(\sigma_y\), while the angular spread can be in the order of 2–5 mm-mrad for a cyclotron beam. It can also be parameterized using the emittance of the beam, defined as the product of the size and angular divergence of the beam in a plane perpendicular to the beam direction. The significance of the beam spot size might depend on the design of the treatment head, here in particular the design of the modulator wheel, as a bigger spot size might impact the number of absorber steps covered by the beam at a given time. The energy and energy spread can be obtained with sufficient accuracy by measuring range and shape of Bragg peaks in water (Bednarz et al., 2011) or using an elastic scattering technique (Brooks et al., 1997). Their values might influence the flatness of an SOBP because of the peak-to-plateau ratio of the individual Bragg peaks that form an SOBP. The energy spread of proton beams entering the treatment head from a cyclotron is typically <1% \((\text{DE}/E)\) while a synchrotron may extract beams with an energy spread two orders of magnitude lower. Clinically most important is certainly the beam energy as fields are prescribed by the beam range in water.

The parameters describing the beam are typically correlated. Such a correlation, for example, between the particle’s position within a beam spot and its angular momentum, needs to be taken into account when modeling proton beam scanning (Dowdell et al., 2012). Consequently, a parameterization of the phase space at treatment head entrance is needed and might be defined based on the knowledge of the magnetic beam steering system, that is, from first principles, or by fitting measured data (Paganetti et al., 2008). Depending on the beam steering, a scanned proton beam can be convergent or divergent. For passive scattered delivery, any correlation will most likely be blurred by the scattering material in the treatment head (Paganetti et al., 2004a). For the same reason, the angular spread at treatment head entrance has little influence on the beam exiting the treatment head in passive scattering.
14.2.2 Monte Carlo Modeling of a Therapy Treatment Head

If we neglect hybrid methods, there are three different types of treatment head concepts in proton therapy. There are treatment heads to deliver beam scanning, passive scattering, or are capable of doing both. Down to which accuracy one needs to model the treatment head in a Monte Carlo system depends on the purpose of the simulation. If the purpose is to calculate phase space distributions for dose calculation in the patient, it is most likely sufficient to have only beam-shaping devices included in the simulation. For beam scanning, these are the scanning magnets plus (if used) a patient-specific aperture. For passive scattering, these include the double- or single-scattering system, the modulator wheel, aperture, and compensator. If dose calculation is not the only purpose of the treatment head model, other devices may have to be included as well, for example, ionization chambers for detector studies or absolute dosimetry, or housing of devices to study scattering or shielding effects. To simulate a realistic dose distribution, simulating treatment head ionization chambers in detail might not be necessary because they cause negligible scattering and energy loss of the beam. Simulating plain or segmented ionization chambers is done for the purpose of designing ionization chambers or studying beam steering, as well as calculating the absolute dose in machine monitor units (Herault et al., 2005, 2007; Paganetti, 2006; Koch et al., 2008). These considerations do not include any devices, for example, detectors, one might want to model for research purposes. Monte Carlo simulations have been performed to design a prompt gamma detector for quality assurance (Kang and Kim, 2009). Another example is the use of Monte Carlo simulations to optimize image reconstruction for proton CT (Li et al., 2006; Schulte et al., 2008). Monte Carlo simulations have also been used to study aperture scattering (van Luijk et al., 2001; Kimstrand et al., 2008b; Titt et al., 2008a) or multileaf collimators to replace patient-specific apertures (Bues et al., 2005).

There are various reports on Monte Carlo treatment head simulations (Paganetti et al., 2004a; Fontenot et al., 2005; Herault et al., 2005, 2007; Newhauser et al., 2005; Polf and Newhauser, 2005; Koch et al., 2008; Titt et al., 2008b; Peterson et al., 2009b; Stankovskiy et al., 2009). Monte Carlo simulations have been used when commissioning planning systems (Koch and Newhauser, 2005; Newhauser et al., 2005, 2007). Furthermore, simulations are very helpful for designing a new facility, for quality assurance, and for supporting everyday operation, for example, to calculate tolerances on the appropriate operational parameters for beam delivery.

Quality assurance in radiation therapy is based on well-defined experimental studies that are repeated frequently. Nevertheless, Monte Carlo simulations can assist clinical quality assurance procedures. By simulating dose distributions and varying beam input parameters, tolerance levels for beam parameters can be defined (Paganetti et al., 2004a). Further, slight uncertainties or misalignments in the treatment head geometry that might occur over time can be simulated.

A treatment planning system defines a field either by specifying the treatment head parameters directly or by specifying the range and modulation width, which are then converted into the treatment head parameters internally by a treatment control system. Other than in photon therapy (where this is true only for collimators and jaws), proton therapy treatment heads are highly patient field specific. For each patient field, there can be a different beam energy (or multiple energies) at treatment head entrance and there are unique settings of devices in the treatment head. For beam scanning, these unique settings might be the setting of the magnetic fields and the scanning pattern, while for passive scattered delivery this refers to settings of the scattering system, choice of the modulator wheel, settings of variable jaws, and patient-specific aperture and compensators. The latter two devices, being mounted at the nozzle exit, might be excluded from the treatment head modeling and incorporated in the patient dose calculation because they are fabricated patient specifically. Typically, at the start of the simulation, a generic treatment head geometry is initialized and then modified using parameters provided via an input file.

The accuracy with which the treatment head elements can be modeled depends on the available information, for example, whether drawings of geometries are provided by the vendor (Paganetti et al., 2004a). In addition, it might depend on our knowledge of exact material compositions and material properties (Bednarz et al., 2011). As in photon Monte Carlo, the modeling of machine-specific components in the treatment head can be done using the manufacturer’s blueprints (Paganetti et al., 2004a, 2008; Titt et al., 2008b). Some of the devices can be modeled relatively easy within a Monte Carlo code as the geometry represents a combination of regular geometrical objects. This holds, for example, in the case of ionization chambers or scattering foils. Irregular-shaped objects can be contoured scatterers (Figure 14.2), modulator wheels (Figure 14.3), as well as apertures and compensators. These can be modeled easily if the Monte Carlo code is capable of reading computer-aided design format or if the Monte Carlo allows the definition of objects by importing boundary points in other formats. If these options do not exist, devices have to be programmed by combining regular-shaped objects.

Contoured scatterers consist of two components, one made out of a high-Z and one made out of a low-Z material. They might be modeled by combining cones (Paganetti et al., 2004a). The thickness of the high-Z material decreases radially with distance to the field center while the thickness of the low-Z material increases as a function of radial distance. The scatterer produces a parallel beam and the bimaterial design ensures that the scattering power is independent of the beam energy.

To simulate a modulator wheel, the wheel steps can be modeled as segments out of a circular structure. Each of these segments can then be characterized by thickness, material, minimum and maximum radius, as well as a start and a stop angle. The rotational motion of the modulator can be considered either by adding a large amount of individual Monte Carlo runs based on distinct geometries, or by changing the geometry dynamically,
applying a 4D Monte Carlo technique (Paganetti, 2004). Because beam spots often overlap with several wheel steps, it is not sufficient to use one simulation per absorber step. For example, in one simulation of passive scattered delivery, the wheel position was changed in steps of 0.7° (Paganetti et al., 2004a).

Because each patient field is unique in terms of range and modulation width, there is a unique wheel design for each field. However, it is impractical to fabricate a modulator wheel for each field. Consequently, one operates with a small set of modulator wheels (or a set of wheels with a set of tracks where each track resembles a unique step design; see Figure 14.3). Each wheel (or track) is designed to create full modulation to the patient’s skin. The desired modulation width is achieved by translating the modulation width into desired rotation angles where the beam current is turned on and off at the cyclotron (beam source) level (Lu and Kooy, 2006). Furthermore, for a

FIGURE 14.2  Monte Carlo simulation (based on the Geant4 code) of one of the contoured scatterers used in a double-scattering system at Massachusetts General Hospital, Boston. Upper left: Schematic drawing. Lower left: Full Monte Carlo representation. Lower right: Subvolume for illustration.

FIGURE 14.3  Left: Picture of one of the modulator wheels consisting of three tracks used at Massachusetts General Hospital, Boston. Right: Simulation of a modulator wheel using the Geant4 Monte Carlo code.
given wheel (track), a flat SOBP is only achieved for a nominal range and modulation width combination. To fine-tune the shape of the SOBP depth dose distribution, the beam current is continuously modulated as a function of rotation angle to ensure that each SOBP satisfies its flatness specification.

Field-specific apertures and compensators are prescribed by the planning system. They are defined in files that parameterize (e.g., by a set of points in space) the geometry for use in a milling machine. How the geometries are imported into the Monte Carlo code depends on the format of the files as well as on the capabilities of the Monte Carlo system. One solution is to use the milling machine files directly. An aperture can be described by a set of points following the inner shape of the aperture opening. If the Monte Carlo code is capable of translating this information into a three-dimensional (3D) object representing the aperture, the information can be used directly. Otherwise, different parameterizations need to be applied, as, for example, a triangulation of the aperture opening (Paganetti et al., 2004a).

For the compensator, one can virtually drill holes into a plastic tube to generate the geometry in the Monte Carlo mimicking the actual device. Apertures need to be modeled explicitly because of the secondary radiation they produce (Zacharatou Jarlskog et al., 2008) and the effects of edge scattering (Titt et al., 2008b).

Some Monte Carlo codes are also capable of simulating magnetic beam steering. Treatment head simulations for scanned beams have shown good agreement (Peterson et al., 2009a,b). Depending on the scanning system, accurate description of the field lines as compared to binary fields, and detailed information of the beam emittance are important when simulating dose distributions using Monte Carlo. To simulate pencil beam scanning, the magnetic field settings are prescribed either by the planning system directly or by a treatment control system based on information by the planning system (Paganetti et al., 2005). These settings are typically relative numbers that need to be translated into magnetic strengths in Tesla. The relationship is typically known from results of commissioning measurements. If the Monte Carlo code is capable of modeling magnetic fields, this information can be incorporated by defining the field strengths as a function of position in the treatment head geometry (Paganetti et al., 2004a). This requires a parameterization of the magnetic field lines. One can also use an approximation by defining a constant magnetic field as either on or off in a defined area (i.e., assume perfect dipoles). Attention needs to be paid to the step size while tracking protons through the magnetic field because the field is applied on a step-by-step basis. Large steps lead to considerable uncertainties when simulating the curved path of particles through the field. Figure 14.4 illustrates the simulation of pencil beam scanning where proton pencils have been steered through the treatment head. To simulate an entire scanning pattern, 4D Monte Carlo techniques can be applied (Paganetti, 2004; Paganetti et al., 2005).

As discussed above, for passive scattering as well as beam scanning, the treatment head geometry involves devices with a time-dependent setting. The rotation of the modulator wheel can be simulated by adding individual runs of pristine Bragg peaks and combining them based on weighting factors obtained from the wheel geometry (Herault et al., 2005; Newhauser et al., 2005). One can also use a slab of a fixed thickness and then change the start position of the beam inside the slab to modulate different thicknesses (Koch et al., 2008). An alternative is to change the geometry dynamically by applying a 4D Monte Carlo technique (Paganetti, 2004; Shin et al., 2012). Furthermore, a time-dependent feature might be the modulation of the beam current in passive scattering systems (Lu and Kooy, 2006). The beam current is continuously modulated as a function of rotation angle. This can be incorporated into the Monte Carlo code by using a finite set of look-up tables correlated to the rotating wheel. To simulate a beam scanning pattern, 4D Monte Carlo can be used to constantly update the magnetic field strength (Paganetti, 2004; Paganetti et al., 2005). This allows studying beam scanning delivery parameters (Paganetti et al., 2005; Peterson et al., 2009a,b; Shin et al., 2012).

14.2.3 Phase Space Distributions

Phase space distributions are commonly used to characterize the beam at or near the treatment head exit. The aim is to minimize calculation time by reusing data in areas that are not varying for each treatment field. In passive scattered proton therapy, phase spaces are less useful than they are in photon therapy. This is because each field typically has a unique setting of the treatment head. The varying parameters are (at least) beam energy, the settings for first scatterer, second scatterer, modulator wheel, utilized wheel solid angle, as well as aperture and compensator.
The same holds for beam scanning as the scanning pattern depends on the treatment field. However, for pencil beam scanning, the definition of a phase space to characterize the radiation field exiting the treatment head is feasible as the variations in beam settings are much lower (beam energy and magnet settings). To scan a particular pattern within the patient, the scanning trajectory for a given energy layer can be discretized. These beam spots, corresponding to magnet settings, can be sampled in a phase space file. For treatment simulation, a predefined number of protons is then simulated for each beam spot.

14.2.4 Beam Models

Beam models allow parameterization of a radiation field to be used instead of a phase space. A model would thus have to incorporate the variety of treatment head settings. Proton therapy treatment head settings are very complex and an accurate beam model might be hard to find. Assuming a homogeneous field impinging on the patient-specific aperture and compensator and making a rough assumption about the angular spread of the field, it is possible to deconvolve an SOBP into its pristine peak contributions (Lu and Kooy, 2006). This method has been suggested for optimizing beam current modulation but could also serve to construct a beam model by optimizing a set of pristine Bragg curves that add up to an SOBP. It is unclear whether such a parameterization can serve as a beam model for patient-related dose calculations because the creation of a given SOBP might not be unique in terms of the underlying pristine Bragg peaks as dose can be parameterized as fluence times linear energy transfer. Thus, there could be more than one solution within certain dose constraints.

Beam models are, however, feasible for pencil beam scanning where a field can be characterized by a fluence map of pencil beams (x, y, beam energy, weight, divergence, and angular spread). Thus, a beam source model can be based on parameterized particle sources for a scanned proton beam using fits of measured data, for example, fluence distributions of pencil beams in air and depth dose distribution measured in water (Kimstrand et al., 2007). The fact that beam models or phase space distributions can easily be constructed for pencil beam scanning has important implications when using Monte Carlo for clinical dose calculation. Owing to the low efficiency of proton therapy treatment heads for passive scattering (typically between 2% and 30%, depending on the field), the majority of calculation time is spent tracking particles through the treatment head. While this makes Monte Carlo for passive scattered delivery less attractive for routine clinical use, the time frame for beam scanning simulations allows the use of Monte Carlo routinely in the clinic (Paganetti et al., 2008). Note that for some scanned beam deliveries, one might want to use an aperture to reduce the beam penumbra. Edge scattering at the aperture can have significant effects on the dose distribution (Titt et al., 2008a) and should thus be included in the beam model.

14.2.5 Uncertainties and Validation

There are various studies comparing experimental depth dose distributions and beam profiles with results from different Monte Carlo codes and for different treatment heads (Paganetti et al., 2004a; Titt et al., 2008b; Clasie et al., 2010). In a well-designed Monte Carlo code, measured dose distributions of an SOBP in water should be reproduced with accuracies within typically ~1 mm in range and ~3 mm in modulation width (Paganetti et al., 2004a). A benchmark example using an inhomogeneous phantom setup is shown in Figure 14.5 with a comparison between measured (ionization chamber) and Monte Carlo predicted dose profile.

Even if an appropriate setting of the physics parameters in the Monte Carlo has been identified, there are still uncertainties because various parameters are not necessarily known down to a sufficient level of accuracy. The outcome of simulations depends, for example, on the known geometry of the treatment nozzle, such as modulator wheel (including the time-dependent settings during particle beam delivery and beam current modulation (Paganetti et al., 2004a)), scatterers, collimators, and patient-specific components (aperture and range compensator).

It has been shown (Bednarz et al., 2011) that the SOBP range is most sensitive to changes in the density of materials in the modulator wheel. Materials commonly used are polyethylene (lexan) and lead. A 10% variation in density could result in range changes on the order of 1 mm. Another material used in scatterers of treatment heads is carbon. Here, the nominal density as specified by the manufacturer can vary substantially from the actual density because carbon is available in various specifications leading to larger uncertainties compared to other materials. This can have an impact on the predicted proton beam range as well.

The uncertainties in the Monte Carlo settings affect not only the beam range but also the modulation width. The flatness of the SOBP might be sensitive to the initial energy spread and spot size at nozzle entrance. These parameters influence the peak-to-plateau ratio of the individual Bragg peaks that form an SOBP. It has been shown that the effect is small for at least certain modulator wheel designs (Bednarz et al., 2011). The impact of the beam spot size depends on the width of each step used on the modulator wheel. Uncertainties in stopping power parameters, especially the mean excitation energy (see above), can also affect dose calculations (Andreo, 2009).

Commissioning of a Monte Carlo treatment head implementation might not be doable based on first principles depending on the complexity of the treatment head or beam settings or the geometry information available. Thus, although good agreement might be found based on first principles (Paganetti et al., 2004a; Bednarz et al., 2011), commissioning of a Monte Carlo-based passive scattering proton beam system might require fine-tuning to ensure accurate dose distributions in the patient. If the simulations differ from experimental data, adjusting parameters can be cumbersome. For example, adjusting nozzle geometries to improve a slight tilt in an SOBP will
correspondingly affect the modulation width. Beam current modulation (beam weight as a function of modulator wheel angle) can be useful to correct the Monte Carlo simulations if the desired SOBP flatness cannot be reached based on first principles (Bednarz et al., 2011).

While the comparison of dose distributions is a valuable benchmark for electromagnetic interactions, nuclear interaction models cannot be validated measuring dose alone. A device to study total inelastic cross sections is the multilayer Faraday cup (Gottschalk et al., 1999; Paganetti and Gottschalk, 2003). It measures the longitudinal charge distribution of primary and secondary particles and is capable of separating the nuclear interaction component from the electromagnetic component. Various Monte Carlo physics models have been compared with the measured charge distribution from the Faraday cup (Zacharatou Jarlskog and Paganetti, 2008a; Rinaldi et al., 2011).

A precise modeling of neutron yields is needed when simulating both neutron doses (Zacharatou Jarlskog et al., 2008; Zacharatou Jarlskog and Paganetti, 2008b) or neutron fluence for shielding. These simulations require double differential production cross sections for tissues, beam-shaping devices, and shielding materials. There are often insufficient experimental data to predict inelastic nuclear cross sections. The agreement between experiment and simulation for neutron doses is typically not as good as with photons, electrons, or protons (Schneider et al., 2002; Polf and Newhauser, 2005; Moyers et al., 2008; Zheng et al., 2008; Chen and Ahmad, 2009; Clasie et al., 2010).

14.3 Simulating Dose to the Patient

14.3.1 Statistical Accuracy of Dose Modeling

Typically, one tries to achieve a statistical uncertainty for dose calculations of less than 2%, at least for the target volume. However, the simulation conditions affecting the uncertainties are slightly different in proton therapy compared to photon therapy. The energy deposited in a patient voxel per proton is higher than in photon Monte Carlo. Thus, typically fewer protons have to be tracked to achieve a targeted accuracy as compared to photon Monte Carlo.

The required number of particle histories depends on the efficiency of the treatment head (determined, for example, by the double-scattering system in passive scattered proton therapy), the required geometrical resolution, and slightly on beam range and modulation width. It normally does not depend on the field size downstream of the patient field-specific aperture because the field size generated in the double-scattering system is...
typically independent of the aperture opening. For pencil beam scanning, the efficiency of the treatment head is not an issue as it will be almost 100%, while for passive scattering it is typically in the order of 3–30%. For passive scattered delivery, a typical number of proton histories for a given patient field is ~25 million at treatment head entrance if sufficient statistical accuracy is required for a single field based on the treatment planning grid resolution. The total number of protons entering the patient for simulating dose distributions can be even less than 1 Mio. for certain fields. These numbers are quite low compared to photon therapy. Owing to the proton linear energy transfer (LET), only a few hundred protons are traversing a voxel for a 2% dose variation. Nevertheless, simulations might still be time consuming and several techniques have been presented to improve the computational efficiency of proton beams in patients (Jiang and Paganetti, 2004; Yepes et al., 2009).

14.3.2 CT Conversion

For photon or electron beams, electron density is being used in analytical dose calculation engines because the dominant energy loss process is interaction with electrons. Protons lose energy by ionizations, multiple coulomb scattering, and non-elastic nuclear reactions. Each interaction type has a different relationship with the material characteristics obtained from the CT scan (Matsufuji et al., 1998; Palmans and Verhaegen, 2005). Consequently, for analytical proton dose calculations, instead of electron density or mass density, relative stopping power is being used to define water-equivalent tissue properties and a conversion from CT numbers to relative stopping power is being used in commercial planning systems (Mustafa and Jackson, 1983; Schaffner and Pedroni, 1998). The accuracy of dose calculations is affected significantly by the ability to precisely define tissues based on CT scans (Schaffner and Pedroni, 1998; Jiang et al., 2007).

For Monte Carlo dose calculation, a conversion from CT numbers to material compositions and mass densities is done for each tissue. A conversion scheme to relate CT numbers to human tissues can be deduced by scanning tissue phantom materials or animal tissues. Note that a CT number reflects the attenuation coefficient of human tissues to diagnostic x-rays and may be identical for several combinations of elemental compositions, elemental weights, and mass densities (Schneider et al., 2000). In CT conversion schemes, tissues are grouped into different tissues sharing the same material composition (and mean excitation energy). The number of groups certainly affects the dose calculation accuracy and is typical between 5 and 30 (Jiang and Paganetti, 2004; Jiang et al., 2007; Parodi et al., 2007a). While grouping material compositions is justified, it is typically not sufficient to assign a unique density to each group, that is, the number of densities is typically the same as the number of gray values (CT numbers) (Jiang and Paganetti, 2004). Several tissue classifications for CT-based Monte Carlo dose calculation have been introduced (Schneider et al., 1996, 2000; du Plessis et al., 1998). Most methods are based on a stoichiometric calibration of CT numbers, which provides a robust division of most soft tissues and skeletal tissues.

Not only can the absolute dose vary but the proton beam range might also depend on the accuracy of the CT conversion (Chvetsov and Paige, 2010; Espana Palomares and Paganetti, 2010). Range uncertainties due to uncertainties in CT conversion are of concern in proton therapy (Unkelbach et al., 2007; Paganetti, 2012). It has been shown that slight discrepancies in mass density assignments play only a minor role in the target region whereas more significant effects are caused by different assignments in elemental compositions (Jiang et al., 2007). The relationship between CT number and material is not unique and various fits can lead to a feasible result (Schneider et al., 2000; Espana Palomares and Paganetti, 2010).

A comparison of two different CT conversion methods (Schneider et al., 2000; Rogers et al., 2002) is shown in Figure 14.6 to illustrate the magnitude of differences one might expect. The largest difference appears at $H = -119$ for carbon and oxygen because of the defined thresholds at the transition of air and adipose tissue. For hydrogen and calcium, the greatest difference is located at $H = +126$ at the transition from soft tissue to bone. To demonstrate the impact these differences might have on proton dose calculation, Figure 14.7 shows the mass stopping power ratio as a function of CT number for the two methods presented in Figure 14.6.

When implementing a conversion scheme for Monte Carlo dose calculation, a normalization to the specific scanner that is being used for imaging in the department is needed to compare the results from a Monte Carlo dose calculation with those from an analytical algorithm in a treatment planning system. The normalization for the Monte Carlo can be found by either doing a separate stoichiometric calibration or, as an approximation, by simulating relative stopping power values based on an existing CT conversion and then compare the results with the planning system conversion curve. Normalization can then be achieved within the Monte Carlo algorithm by slightly adjusting material compositions or, easier, mass densities (Jiang and Paganetti, 2004; Parodi et al., 2007b; Paganetti et al., 2008).

Uncertainties in dose calculation are introduced not only by CT conversion but also because of the imaging resolution, that is, the CT grid size (Chvetsov and Paige, 2010; Espana Palomares and Paganetti, 2010). Treatment planning programs often report dose on a more coarse grid than the one provided by the patient’s CT scan. The Monte Carlo simulation should operate on the actual CT scan resolution, which is typically on the order of 0.5–5 mm. It happens frequently that the CT grid used clinically is nonuniform. For comparison of the results with the planning system (based on the planning grid), the Monte Carlo dose distribution can be interpolated. The speed of the Monte Carlo simulation certainly depends on the grid size. However, assuming that a step size above 1 mm is typically not warranted, a larger grid size does not translate into a huge gain.
14.3.3 Absolute Dose

The prescribed dose is converted into machine monitor units for patient treatment, that is, absolute doses are reported as cGy per MU. In proton therapy, this is done either by calibration measurements, by using analytical algorithms or by relying on empirical data. A monitor unit typically corresponds to a fixed amount of charge collected in a transmission ionization chamber incorporated in the treatment head whose reading is related to a dose at a reference point in water (Kooy et al., 2003, 2005). Using Monte Carlo, one can actually simulate the ionization chamber output charge when tracking particles through the treatment nozzle. Here, an exact model of the proton beam delivery nozzle, including devices used for dosimetry, is required (Paganetti, 2006). The disadvantage of this method is that it typically requires a large number of histories to be simulated as the energy deposited in a small reference chambers in air per particle is normally quite small. Nevertheless, Monte Carlo predicted output factors can be part of an extended quality assurance program because the influence of specific devices in the treatment head on the output factor reading can be studied.

Alternatively, absolute dosimetry for Monte Carlo dose calculation can also be done by simply relating the number of protons...
at nozzle entrance to the dose in an SOBP in water. With an accurate model of the treatment head, this method is equivalent to a direct output simulation because instead of indirectly “measuring” the number of protons at a given plane in the treatment head (in an ionization chamber), one “measures” the number of protons at nozzle entrance.

14.3.4 Dose-to-Water and Dose-to-Tissue

Traditionally, dose in radiation therapy is reported as dose-to-water because analytical dose calculation engines, as the ones incorporated in commercial planning systems, calculate dose by modeling physics relative to water. Furthermore, quality assurance and absolute dose measurements are done in water. There is an ongoing debate about the pros and cons of using the two standards (Liu and Keall, 2002). Monte Carlo dose calculation systems calculate dose-to-tissue because they allow the material properties, as converted from CT numbers, to be modeled using explicit material composition, mass density, and excitation energy. Consequently, to allow a proper comparison between Monte Carlo and pencil beam-generated dose distributions, one has to convert one dose metric to the other.

Based on relative stopping powers and on a nuclear interaction parameterization, a formalism has been described to convert dose-to-medium into dose-to-water for proton beam dose calculations (Paganetti, 2009a). It was found that based on the mean dose to a structure, dose-to-water can be higher by ~10–15% compared to dose-to-tissue in bony anatomy. For soft tissues, the differences were shown to be on the order of 2%. The difference in mean dose roughly scales linear with the average CT number, and thus seems to be clinically insignificant for soft tissues but needs to be taken into account for bony anatomy. It has been shown that in most cases it is sufficiently accurate (within ~1%) to do a conversion to dose-to-water retroactively by simply multiplying the dose with energy-independent relative stopping powers. Furthermore, as the difference in mean dose roughly scales linear with the average CT number, a rough scaling based on the CT numbers might be sufficient, depending on the desired precision (Paganetti, 2009a).

14.3.5 Impact of Nuclear Interaction Products on Patient Dose Distributions

Pencil beam algorithms in treatment planning systems incorporate nuclear interaction contributions intrinsically because they are based on measured or Monte Carlo simulated depth dose curves. It has been shown that nuclear secondaries need to be considered in proton therapy dose calculations as they contribute significantly in particular in the entrance region of the Bragg curve (Carlsson and Carlsson, 1977; Laitano et al., 1996; Medin and Andreo, 1997; Paganetti, 2002; Palmans and Verhaegen, 2005).

Depending on the application, it might not be necessary to track all secondary particles in Monte Carlo dose calculation and it might be sufficient for some types to deposit the energy locally (to ensure proper energy balance). If secondary doses for out-of-field dose calculations are being studied, one also has to consider neutrons and the protons they cause. For patient-related dose calculations, secondary protons from proton–nuclear interactions solely need to be tracked. Secondary protons deliver up to about 10% of the total dose proximal to the Bragg peak of an unmodulated proton beam. Furthermore, secondary protons cause a dose build-up due to forward emission of secondary particles from nuclear interactions. Because the total nuclear interaction cross section shows a maximum at around 20 MeV and decreases sharply if the energy is decreased, the contribution of dose due to nuclear interactions becomes negligible roughly at the Bragg peak position of a pristine Bragg curve. As a rule of thumb, the average proton energy in the Bragg peak is ~10% of the initial energy. In an SOBP, nuclear interactions can still play a role because dose regions proximal to the Bragg peak also contribute and thus cause a tilt of the dose plateau if their contribution is neglected (Paganetti, 2002). Figure 14.8 demonstrates the magnitude of the dose from secondary protons as well as its 3D distribution in a patient.

Note that the consideration of nuclear interactions is particularly important for pencil beam scanning since each pencil is surrounded by a nuclear halo from secondary protons. Neglecting it will result in significant deviations in absolute dose (Sawakuchi et al., 2008, 2010).

14.3.6 Differences between Proton Monte Carlo and Pencil Beam Dose Calculation

It has been shown that the dosimetric differences between a Monte Carlo and a pencil beam algorithm can be significant (Petti, 1996; Schaffner et al., 1999; Szymanowski and Oelfke, 2002; Soukup et al., 2005; Tourovsky et al., 2005; Soukup and Alber, 2007; Paganetti et al., 2008). In fact, the routine use of Monte Carlo dose calculation in proton therapy could result in a reduction of treatment planning margins, which in turn might reduce treatment side effects (Paganetti, 2012). Monte Carlo simulations have been used to validate analytical algorithms (Sandison et al., 1997; Soukup et al., 2005; Paganetti et al., 2008) or for commissioning of treatment planning systems (Koch and Newhauser, 2005; Newhauser et al., 2007). Overall, some pencil beam algorithms show good agreement with Monte Carlo (Paganetti et al., 2008). Pencil beam algorithms are less sensitive to complex geometries and in-beam density variations, that is, bone–soft tissue, bone–air, or air–soft tissue interfaces (Urie et al., 1984; Petti, 1992, 1996; Pflugfelder et al., 2007). Deviations from the Monte Carlo dose distribution occur specifically in the penumbra if the beam direction is tangential to an interface connecting high and low CT numbers. Not only local difference can occur. Also the overall dose distribution can be affected. If proton beams pass through complex heterogeneous geometries, a phenomenon called range degradation occurs (Urie et al., 1986; Sawakuchi et al., 2008), which is correctly predicted only when using...
Monte Carlo. Note also that some pencil beam models do not consider aperture edge scattering (Titt et al., 2008b).

Consequently, one of the areas where pencil beam algorithms specifically show weaknesses is in the presence of lateral heterogeneities (Schaffner et al., 1999; Soukup et al., 2005; Soukup and Alber, 2007). Figure 14.9 demonstrates one of the weaknesses of a pencil beam-generated dose distribution, namely, the inaccurate consideration of interfaces parallel to the beam path. This patient was treated for a spinal cord astrocytoma with three coplanar fields. The figure only shows one of the fields. The Monte Carlo dose calculation was based on a CT with 176 × 147 × 126 slices with voxel dimensions of 0.932 × 0.932 × 2.5–3.75 mm³ (slice thickness varied).

A treatment technique that is being used in passive scattered proton therapy is field patching. Here, a lateral penumbra is matched with a distal fall-off of a second beam to produce an L-shaped dose distribution with just two beams. This places the distal fall-off of one of the beams in the target, making this technique vulnerable to range uncertainties. To minimize the effect of range uncertainties, multiple patch field combinations are being applied. Further, the treatment planners tend to overshoot the field with the distal fall-off ending in the target to avoid a hot spot. Figure 14.10 shows an example for a patch field combination and the impact of differences in the predicted range when using pencil beam or Monte Carlo-based dose calculation. The Monte Carlo dose calculation was based on a CT with 161 × 177 × 111

**FIGURE 14.8** (See color insert.) Impact of nuclear secondaries on the dose distribution in a patient. Left: Dose distribution for one field (out of three for this treatment plan) for a head and neck case (doses in%). Right: Contribution of secondary protons to the dose distribution in% of the target dose.

**FIGURE 14.9** (See color insert.) Axial views of dose distributions calculated using Monte Carlo (left) and a commercial planning system based on a pencil beam algorithm (right). Doses are in%. The red dashed circles show the dosimetric impact of an interface parallel to the beam path. (From Paganetti, H. et al., Clinical implementation of full Monte Carlo dose calculation in proton beam therapy. *Physics in Medicine and Biology*, 2008. 53(17): 4825–4853. With permission.)
slices with voxel dimensions of $0.65 \times 0.65 \times 1.25–3.75$ mm$^3$ (variable slice thickness). The patient had a sphenoid sinus tumor. One can see nicely the differences between pencil beam and Monte Carlo at the end of range, that is, at the patch lines. In this case, the consequence is a hot spot in the target. Also apparent is a discrepancy in the penumbra of the field shown on the right. The reason might be the sharp gradient in the range compensator not modeled exactly in the pencil beam algorithm.

Whether Monte Carlo dose calculation is needed or pencil beam-generated dose distributions are sufficiently accurate is still an area of research. Treatment planners are typically aware of dose calculation uncertainties and take these into account by applying appropriate margins. Furthermore, one avoids pointing the beam toward a critical structure because of uncertainties in the position and gradient of the distal dose fall-off. Because of conservative margins in treatment planning, differences between analytical and Monte Carlo dose calculations are not always clinically significant (Soukup and Alber, 2007; Paganetti et al., 2008). However, more accurate dose calculation might help reducing those margin contributions that are due to dose calculation uncertainties (Paganetti, 2012).

Monte Carlo dose calculations are attractive for studying motion effects in lung dosimetry (Keall et al., 2004; Paganetti et al., 2004b). In proton therapy, 4D Monte Carlo simulations have been used to model respiratory patient motion (Paganetti, 2004; Paganetti et al., 2004b) and to study interplay effects.
between respiratory motion and beam scanning speed (Paganetti et al., 2005).

Obviously, Monte Carlo dose calculations are particularly valuable for studying dosimetric effects in regions with a high gradient in density. Extreme cases are metallic implants in the patient, which are being used to stabilize bony anatomy after surgery or as markers for imaging. Monte Carlo has also been used to study Tantalum markers used to stabilize bony anatomy after surgery or as markers for imaging. These can lead to significant dose perturbations typically not predicted accurately by pencil beam algorithms (Newhauser et al., 2007). Markers implanted in the patient for setup or motion tracking are typically not modeled accurately in pencil beam algorithms due to their high-Z nature.

14.3.7 Clinical Implementation

If Monte Carlo treatment planning is to be done, it needs to be incorporated into a framework for treatment plan optimization (Moravek et al., 2009). Full Monte Carlo treatment planning is time consuming because the dose would have to be simulated multiple times during the optimization process. Alternatively, one might utilize a Monte Carlo dose engine only at a limited number of checkpoints during the optimization process. Monte Carlo can also be used outside of a commercial planning system as a standalone tool to get a second opinion on a dose distribution predicted by an analytical algorithm.

For an in-house-developed Monte Carlo dose calculation engine, treatment head simulation and the capability of the Monte Carlo code to simulate dose on a CT grid are the only step toward clinical implementation. The Monte Carlo simulation has to be based on treatment plan parameters, which have to be imported. How these parameters are extracted depends on the treatment planning program and on the access it provides to those data.

If the radiation field for proton beam scanning is to be simulated, the planning system will most likely provide a matrix of beam spot energy, beam spot weight, and beam spot position, which can be translated into Monte Carlo settings in a straightforward manner. If passive scattered delivery is simulated, there are two ways of how the radiation field is characterized. First, the planning system prescribes the settings of various treatment head devices and the beam at treatment head entrance directly. These settings are beam energy at nozzle entrance, settings of the first and second scatterer (one scatterer if single scattering is used for small fields), and a combination of scatterer steps in a modulator wheel or a ridge filter. Second, the planning system prescribes range and modulation width only, leaving it to the hardware control system of the facility to translate this information into treatment head settings. In this case, it might be warranted to incorporate the control system algorithms into the Monte Carlo code (Paganetti et al., 2004a, 2008). Planning systems also define aperture and compensator for passive scattered delivery. These are typically given as files readable by a milling machine.

Regarding the patient’s CT, a Monte Carlo code can either accept a digital imaging and communications in medicine (DICOM) stream directly (Kimura et al., 2004; 2005) or receive the CT information from the planning system (Parodi, K. et al., 2007b; Paganetti et al., 2008). For the latter, the patient geometry is transferred from a departmental patient database to the planning system, which in turn transfers the CT to Monte Carlo using a planning system-specific data format. Within the Monte Carlo, it might be translated into a simple binary format. Parameters required are the number of voxels and the, potentially variable, voxel dimensions in x, y, and z. The latter method is less flexible than using DICOM directly but, depending on the planning system, it might be the only way to allow all simulation-related parameters to be imported from one single source, that is, the planning system.

Additional simulation input parameters needed from the planning system for each field are the gantry angle, the patient couch angle, the air gap between patients and treatment head, the isocenter position, and the prescribed dose or beam weight. Again, to import parameters into the Monte Carlo system, a software link to the treatment planning system needs to be established (Paganetti et al., 2008). Note that the spatial information might be given in different coordinate systems, that is, the treatment head coordinate system, the patient CT coordinate system, or the coordinate system of the planning system.

To facilitate the data flow from the planning system to the Monte Carlo, one might use a user interface where the user selects a specific patient, plan, and treatment field. Once the dose in the patient is calculated, it can be analyzed by a standalone program capable of visualizing CT data and reading/visualizing contours. If the planning system allows, the Monte Carlo results can be imported into the planning system for a side-by-side comparison with the analytical results.

The Francis H Burr Proton Therapy Center at Massachusetts General Hospital uses a standalone script that allows the user to extract treatment-related parameters directly from different planning systems for passive scattered or scanned beam proton therapy and translates the information directly into input parameters for the Monte Carlo system TOPAS (Perl et al., 2012), allowing a convenient interface for clinical personnel to do proton Monte Carlo dose calculation.

14.3.8 Improving Monte Carlo Efficiency

Particle tracking in a voxel geometry can be computationally inefficient. Typically, each particle has to stop when a boundary between two different volumes is crossed. Algorithms have been developed to tackle this problem (Hubert-Tremblay et al., 2006; Sarrut and Guigues, 2008). Such algorithms are based on an image segmentation that compromises the regular voxel geometry. There are several techniques to improve the computational efficiency of Monte Carlo for proton beams in patients (Jiang and Paganetti, 2004; Yepes et al., 2009; Schumann et al., 2012). Some Monte Carlo codes have been specifically designed for
fast patient dose calculations using approximations to improve efficiency (Kohno et al., 2002, 2003; Fippel and Soukup, 2004; Tourovskey et al., 2005). One method is to implement a track repeating algorithm in which precalculated proton tracks and their interactions in material are tabulated and used in dose calculations (Li et al., 2005).

To improve calculation efficiency, hybrid methods have been introduced that combine aspects of Monte Carlo simulation and analytical algorithms. There are many ways to improve the accuracy of pencil beam algorithms by adding Monte Carlo components, for example, using spot decomposition. These are often associated with significantly decreased computational efficiency (Soukup et al., 2005; Soukup and Alber, 2007). One might also use Monte Carlo generated kernels as look-up tables in analytical dose calculation algorithms (Soukup and Alber, 2007; Kimstrand et al., 2008b). Simplified Monte Carlo methods using measured depth dose curves in water or other materials as input have also been proposed (Kohno et al., 2002).

A significant speed improvement compared to codes like Geant4 and FLUKA has been reported for VMCpro (Fippel and Soukup, 2004). Here, various approximations were introduced, for example, a simplified multiple scattering algorithm and density scaling functions instead of actual material compositions. Nuclear interactions are treated as a correction to electromagnetic interactions by using parameterizations and distribution sampling and it was shown that these are valid for dose calculations, at least for broad beams.

14.4 Proton Monte Carlo Applications Other Than Treatment Head Simulations and Dose Calculations

14.4.1 Simulating Proton Induced PET Distributions in the Patient

Protons undergo nuclear interactions in the patient, which can lead to the formation of positron emitters. One can use positron emission tomography (PET) imaging for in vivo verification of treatment delivery and, in particular, beam range. Monte Carlo simulations are typically used to generate a theoretical PET image based on the prescribed radiation field, which can then be compared to the measured PET distribution for treatment verification (Parodi et al., 2007a,b). Lack of detailed cross-section data is currently one of the limiting factors with respect to accuracy of in vivo range verification (España et al., 2011). Similarly, the emission of gamma rays from excited nuclear states, the so-called prompt gammas, can be simulated using Monte Carlo (Moteabbed et al., 2011).

14.4.2 Simulating LET Distributions for Radiobiological Considerations

The majority of clinical data in radiation therapy is based on treatments with photon beams, and prescription doses to cancerous tissue as well as dose constraints to organs at risk are based on clinical experience and not necessarily on explicit radiobiological considerations. Proton doses are prescribed relative to photon doses by applying the concept of relative biological effectiveness (RBE) for proton therapy (Paganetti et al., 2002). The RBE of protons depends on the linear energy transfer (among other parameters like dose, biological endpoint, dose rate, etc.).

Distributions of LET (or better dose-averaged and track-averaged LET) can be simulated in the patient using Monte Carlo (Grassberger and Paganetti, 2011). In the target area, the dose is typically delivered homogeneous. This however does not guarantee a homogeneous distribution of LET values (Paganetti and Schmitz, 1996; Paganetti and Goitein, 2000). When considering primary particles only, elevated LET values will appear in the distal end of a proton field. Furthermore, nuclear interaction products may cause elevated LET values even outside of the target volume, that is, in the entrance region of the Bragg curve (Kempe et al., 2007). One might simulate dose-averaged LET distributions in a patient geometry to identify potential hot spots of biological effectiveness (Grassberger et al., 2011).

14.4.3 Simulating Secondary Neutron Doses to the Patient

Monte Carlo simulations have been used for shielding design studies (Binns and Hough, 1997; Fan et al., 2007; Agosteo, 2009). Secondary neutron doses from proton therapy are difficult to measure because neutrons are indirectly ionizing and interact sparsely, causing only low absorbed doses. The MCNPX code (Agosteo et al., 1998; Newhauser et al., 2005; Polf and Newhauser, 2005; Polf et al., 2005; Tayama et al., 2006; Zheng et al., 2007, 2008; Fontenot et al., 2008; Moyers et al., 2008; Taddei et al., 2008, 2009; Perez-Andujar et al., 2009), FLUKA (Agosteo et al., 1998; Schneider et al., 2002), and GEANT4 (Jiang et al., 2005; Zacharatou Jarlskog et al., 2008; Athar and Paganetti, 2009; Athar et al., 2010) were applied to assess secondary doses in proton beams. Neutron production in proton beams was also studied using the Shield-Hit (Gudowska et al., 2002; Gudowska and Sobolevsky, 2005) and PHITS codes (Sato et al., 2009).

There are considerable uncertainties when it comes to simulating secondary particle production because there is insufficient experimental data of inelastic nuclear cross sections in the energy region of interest. Neutron and secondary charged particle emissions from nuclear interactions can be the result of complex interactions. Fast neutrons lose most of their kinetic energy in the initial relatively small number of interactions. In the low/thermal energy region, there are a large number of elastic scatterings in soft tissues, causing the neutron energy distributions in the patient to be dominated by low-energy neutrons (Jiang et al., 2005).

At low doses, the quantity of interest is not the absorbed dose but the equivalent dose. Calculations of the secondary equivalent doses to patients require particle and particle energy-dependent radiation weighting factors to consider the biological
effectiveness. This is different from RBE as radiation weighting factors are regulatory quantities at low dose (ICRP, 2003a).

One possible strategy is to calculate the average absorbed dose for the organ under consideration and scale the dose at each interaction point with a radiation weighting factor. Another approach frequently used is to calculate the particle fluences at the surface of a region of interest (organ) and then use energy-dependent fluence-to-equivalent dose conversion coefficients (NCRP, 1973; Boag, 1975; ICRU, 1998; Bozkurt et al., 2000, 2001; Chao et al., 2001a,b; Alghamdi et al., 2005; Polf and Newhauser, 2005; Chen, 2006; Zheng et al., 2007).

Whole-body computational phantoms can play an important role when combined with Monte Carlo dose calculations to simulate scattered or secondary doses to organs outside the area imaged for treatment planning (Paganetti, 2009b). Phantoms have been implemented in many Monte Carlo codes to assess neutron doses in proton therapy (Agosteo et al., 1998; Jiang et al., 2005; Zacharatou Jarlsgod et al., 2008; Athar and Paganetti, 2009; Taddei et al., 2009; Athar et al., 2010). Voxel phantoms are largely based on CT images and manually segmented organ contours. For each organ and model, age- and gender-dependent densities, as well as age-dependent material compositions, can be adopted based on ICRU (1992) and organ-specific material composition as a function of age can be based on individuals at the ICRP reference ages (ICRU, 1989; ICRP, 2003b). In addition to standard adult male or female models, models of pregnant patients (Caon et al., 1999; Lee and Bolch, 2003) and the pediatric population have been designed (Zankl et al., 1988; Caon et al., 1999; Nipper et al., 2002; Lee and Bolch, 2003; Staton et al., 2003; Lee et al., 2005, 2006, 2007, 2008). The latest developments in whole-body computational phantoms are hybrid phantoms based on combinations of polygon mesh and nonuniform rational B-spline surfaces (Johnson et al., 2009; Lee et al., 2010). Such phantoms provide the flexibility to model thin tissue layers and allow for free-form phantom deformations for selected body regions and internal organs using patient-specific images and deformable image registration (Tsui et al., 1994; Segars et al., 1999; Segars and Tsui, 2002; Garrity et al., 2003; Lee et al., 2010). Patient body weight can be accommodated through adjustments in adipose tissue distribution (Slyper, 1998; Thirion, 1998; Rueckert et al., 1999). It has been demonstrated that hybrid phantom height matching results in more accurate organ specific neutron dose calculations in proton therapy using Monte Carlo (Moteabbed et al., 2012).

References


Caon, M., G. Bibbo, and J. Pattison, An EGS4-ready tomographic computational model of a 14-year-old female torso for


Chao, T.C., A. Bozkurt, and X.G. Xu, Conversion coefficients based on the VIP-Man anatomical model and GS4-VLSI code for external monoenergetic photons from 10 keV to 10 MeV. *Health Physics*, 2001a. 81: 163–183.


ICRP, Relative Biological Effectiveness (RBE), Quality Factor (Q), and Radiation Weighting Factor (wR). International Commission on Radiological Protection (London, UK: Elsevier Ltd), 2003b. 92.


15

Application of Monte Carlo Methods to Radionuclide Therapy

15.1 Introduction

When treating patients with systemic malignant diseases, external radiation treatment is sometimes not an option because of the wide spread of the tumors. A possible alternative for local treatment with external radiation is therefore the use of radionuclide therapy. In this modality, a radionuclide with a proper decay scheme is carried to the location of the disease by an appropriate radiopharmaceutical or biomolecule and thereby delivers energy by the emission of charged particles to kill the tumor cells. A classic example of radionuclide therapy is the treatment of cancer in the thyroid where orally administered $^{131}$I is taken up naturally by the thyroid as part of its metabolism. If the radionuclide is not active in the metabolism of the tumor/organ, then it can be chemically labeled to a pharmaceutical, antibody, or a peptide. In radionuclide therapy, the major component to the absorbed dose is delivered by radionuclides that emit $\beta$-particles or $\alpha$-particles. However, absorbed dose planning for radionuclide therapy is not a trivial task since the source cannot be turned on–off as is the case in external beam therapy. Instead, the radionuclide decays exponentially with characteristics depending both on the biokinetic properties and on the physical half-life of the radionuclide. On a small-scale level, the radiopharmaceutical is usually heterogeneously distributed, which indicates that the energy deposition may be nonuniform. The biokinetic may also change over time by internal redistribution, which means that activity distribution needs to be measured at several time points to estimate the total number of decays and the related emitted kinetic energy. The dose rate for radionuclide therapy is also much lower than in external beam therapy, which has radiobiological implications. Since radionuclide treatment is systemic, some circulating activity is unavoidable and may result in uptakes in normal organs and tissues. This poses a problem that may restrict the amount of activity to be possibly administered.

Since the biokinetics varies between patients and also over time, the activity uptake and clearance should therefore be measured for each individual patient to estimate the total number of decays in a particular organ/tissue. The way that this can be performed is either by sequential planar scintillation camera measurements or by single-photon emission computed tomography (SPECT) methods. There are several serious problems associated with measuring activity and activity concentrations with a scintillation camera. First, the spatial resolution is relatively poor and in the order of 1–2 cm. This means that small activity volumes cannot be resolved properly due to the partial volume effect. The spatial resolution also varies with the source-to-detector distance. The statistics in the acquired images is a direct function of the activity administered and time for the acquisition and since the sensitivity of the camera is low, statistically related noise will appear in the image. Also, photon attenuation and contribution from scattered photons are effects that need to be compensated for as will be discussed further on in the chapter. Thus, measuring and characterizing the
source distribution in vivo from scintillation camera measurements will be subjected to relatively large uncertainties.

### 15.2 MIRD Formalism

The absorbed dose is defined as the average energy deposit within a mass element. The medical internal radiation dose (MIRD) formalism (Watson et al., 1993; Bolch et al., 2009) has defined the description of internal dosimetry with radionuclides with the following simple equation:

$$D = \tilde{A} \cdot S$$  \hspace{1cm} (15.1)

where $\tilde{A}$ is the cumulative activity (total number of disintegrations during a time interval) and $S$ is a geometry-dependent factor that relates the emitted energy from the source volume to the absorbed energy in the target volume normalized to the mass of the target. The unit of the $S$ value will then be GyMBq⁻¹ h⁻¹. The $S$ value can be described in more detail as

$$S = \frac{\sum n_i E_i \phi(t \leftarrow s)}{m}$$  \hspace{1cm} (15.2)

where $n_i$ is the number of particles that have the energy $E_i$ and the summation is made for all possible particles $i$ in the decay scheme. $\phi$ is the fraction of energy released in the source volume $s$ that will be absorbed in the target volume $t$, and $m$ is the mass of the target volume (usually called absorbed fraction). Since $\phi$ are generally impossible to measure clinically, the only way to obtain relevant data of $\phi$ for a particular geometry is to create a computer phantom of the object (patient) and use this phantom in a full Monte Carlo simulation of the radiation transport to score the energy deposited in the target volume for each source organ. By doing this for each source and target combination, the $S$ values can be tabulated.

### 15.3 Scintillation Camera

As stated in the introduction, the cumulated activity $\tilde{A}$ can be determined using sequential planar scintillation camera measurements or sequential SPECT scans. The scintillation camera consists of a NaI(Tl) crystal so that, when photons are absorbed in the crystal, the energy deposit is converted to visible light through a deexcitation process. This light (proportional to the deposited energy) is guided to several multiplier tubes (PMTs) that each amplifies the weak energy signal, carried by the scintillation light, and generated at the surface of the PMT by a photo cathode. The difference between a counting detector and a camera is that a relevant position of the energy deposition in the crystal can be measured with a good accuracy by calculating the centroid of the measured energy signal obtained from several PMTs. To get an image, however, a collimator is needed. The most commonly used collimator is the parallel-hole collimator, which essentially is a thick sheet of lead and including thousands of hexagonal holes orthogonal to the normal of the crystal. The purpose of the collimator is to absorb all photons except those impinging in a very narrow acceptance angle. The image created will be in two dimensions (2D) with no depth resolution. The spatial resolution is mainly determined by the dimension of the collimator holes and by the distance between the source and the detector and is of the order of 1–1.5 cm. If parallel-hole collimators are used, the sensitivity (cps/MBq) is close to being constant when moving the source within the field of view (FOV). This makes the calibration of measured counts to units of activity feasible for a source in a scatter- and attenuation-free environment. The energy resolution (i.e., the ability to measure energy) is relatively poor for NaI(Tl) crystals equipping conventional gamma cameras and is in the order of 10% full width at half maximum (FWHM) at 140 keV.

### 15.4 Activity Quantitation

From Equation 15.1, it is clear that the cumulated activity within the patient needs to be calculated with a high accuracy since this directly affects the absorbed dose. The activity is commonly measured using a scintillation camera that obtains an image of the activity distribution either in 2D mode (planar) or in 3D mode (SPECT). The second of these two modalities should be preferred because the absorbed energy is actually distributed in 3D.

There are several problems associated with measuring activity with 2D imaging methods, so we will focus on 3D tomographic methods. The SPECT technique is based on projections of events coming from photons acquired in multiple views around the patient. From this set of projections, a set of tomographic images can be reconstructed using a mathematical algorithm. The most commonly used reconstruction method today is the ML-EM (maximum likelihood expectation maximization) iterative method and is based on the Poisson statistics inherent in scintillation camera images. The principle is that the imaging process is modeled by a computer algorithm to mimic a real measurement. This means that projections are calculated from an initial guess of the activity distribution that often is a matrix where all voxels are set to unity. The calculated projections are then compared to real measured projections, and in the ML-EM algorithm, it is the ratio between measured and calculated projection that is used. The ratio is calculated for all projection angles and is then back-projected to obtain a reconstructed error image. The initial image is then multiplied with this error image and normalized and the updated image is then used as an input to the same process. This makes the ML-EM reconstruction to be an iterative process.

The procedure can be described by the following equation:

$$f_j^{k+1} = \frac{f_j^k}{\sum_{i=1}^{n} a_{ij} \cdot \sum_{j=1}^{n} g_i \cdot f_j^k}$$  \hspace{1cm} (15.3)

Here, $a_{ij}$ is the system matrix that describes the probability for a photon originated from a location $j$ to generate an event in the
calculated projection at a location \( i \) in the measured protection \( g \). After performing several iterations, the updated image will converge to a useful image, which means that the numbers in the error image will converge to unity values. The ML-EM method has some attractive mathematical properties but is generally slow in generating an acceptable image since it requires several tens of iterations before convergence. A popular alternative method to the ML-EM method is the OS-EM (ordered-subset expectation-maximization) method (Hudson and Larkin, 1994). This method is very similar to ML-EM but with the important difference in that updates of the activity image are made after a subset of selected projections instead of after all projections have been processed (as is the case for ML-EM).

15.4.1 Photon Attenuation

Photon interactions occur in the patient mainly by photo absorption and Compton and coherent scattering, which means that some photons (otherwise expected to be detected) will be absorbed or scattered in other directions with potentially loss of energy. The scaling from count rate to activity by the use of a single calibration factor \((\text{cps/MBq})\) will then not be valid. The attenuation becomes complex because it also depends on the patient body habitus, the photon energy, and variation in density and tissue composition. In cardiac imaging, nonuniform photon attenuation in regions such as the thorax can introduce artifacts and related false positive defects and therefore correction for this effect is necessary. On modern hybrid SPECT/CT systems, attenuation correction can be made using transmission maps of the patient obtained from a registered CT image. The Hounsfield numbers from the CT study however need to be converted to attenuation coefficients for the relevant radionuclide. Here, validation with Monte Carlo simulations can be of importance. From a digital phantom, created from a physical phantom, a direct comparison between experimental measurements and simulations can be made. The major advantage of simulations is that the interactions in the phantom can be turned off and the result of such a simulation for SPECT application can be a tomographic image representing a scatter- and attenuation-free situation. These images thus can serve as a reference image when evaluating the accuracy and precision in attenuation compensation.

15.4.2 Scatter Contribution to Image

The problem with scatter contribution in SPECT images is perhaps the area where Monte Carlo simulations have been proven to be most useful. The contribution of scatter has its underlying cause in the poor energy resolution of the NaI(Tl) crystal. Therefore, events from scattered photons with lower energy and change in direction will contaminate the image, leading to a reduction in image contrast. Also, here, the scatter contribution depends on the photon energy, source distribution, and on body composition in addition to camera-specific parameters, such as energy window size and location. The Monte Carlo method allows for a separation of the image into components of events from primary unscattered photons, scattered photons, and even further separation of scatter events into categories of scatter orders. Examples of such simulations are shown in Figure 15.1, where the events are separated as a function of the number of scattering interactions they underwent. These curves are

**FIGURE 15.1** Energy pulse-height distribution simulated for \(^{99}\text{Tc}\) and a NaI(Tl) scintillation camera. An energy resolution of 10% (FWHM) at 140 keV has been assumed. The scatter distribution is also shown together with the distribution of first, second, third, or more orders of scattering. Since the spatial distribution of the events in a lower secondary scatter window depends on the location, these types of curves are important to access to optimize the scatter window locations.
impossible to measure experimentally but can provide useful insight of the imaging process.

Many investigators have used such types of Monte Carlo simulation to optimize the locations of additional scatter windows (Holstensson et al., 2007) that are used in energy-window-based scatter compensation methods such as the triple energy window (TEW) method (Ogawa et al., 1991). The TEW method is based on an acquisition in two additional narrow energy windows located in close proximity to the main energy window. The scatter component in the main energy window is then estimated by the average of the images acquired in the two small windows on a pixel-by-pixel basis taking into account the difference in energy window size $W$. This method can be described as

$$I_{\text{scatter corrected}} = I_{\text{main}} - \left( \frac{I_{\text{upper}}}{W_{\text{upper}}} + \frac{I_{\text{lower}}}{W_{\text{lower}}} \right) \cdot \frac{W_{\text{main}}}{2} \quad (15.4)$$

where $I$ are images acquired in the upper, the lower, and the main energy window and $W$ are the corresponding energy window sizes in units of keV. The scatter in the main energy window cannot be measured directly but can be estimated by Monte Carlo methods to be used in a direct comparison between the true scatter distribution and the estimated scatter distribution. An example of such a comparison is shown in Figure 15.2 for $^{99m}\text{Tc}$. The ranges of the lower and upper scatter windows are 120–126 and 154–160 keV, respectively, and the main energy window range is 126–154 keV. A standard Anger-type scintillation camera with a low-energy high-resolution camera and 10% (FWHM) energy resolution and 140 keV was simulated together with the XCAT anthropomorphic phantom (described below in the chapter).

Scatter compensation can also be included in the forward projector of modern iterative reconstruction methods. This can be either a scatter estimation obtained from a TEW acquisition or a model-based scatter distribution. An example of model-based scatter compensation, including Monte Carlo simulated data is the effective scatter source estimator (ESSE) method that was developed by Frey and colleagues (Frey and Tsui 1996). This method provides spatial-domain scatter estimation where the scatter is modeled from an estimate of the activity distribution. The ESSE calculates an effective scatter source from which the attenuated projection of the scatter source will be the estimate of the scatter component in a SPECT projection. The calculation is based on point source scatter kernels, obtained by Monte Carlo simulations for various depths in a water phantom. The method assumes that the photons are propagating through a uniform media (often water) from point source to the last scattering point before the photon exits the phantom toward the camera. Despite this approximation (that has some limitations when the attenuation is nonuniform), the method has shown to be accurate for other isotopes than $^{99m}\text{Tc}$, such as $^{111}\text{In}$ and $^{131}\text{I}$ (Ljungberg et al., 2002, 2003).

Energy-window-based methods are not well suited for bremsstrahlung SPECT of, for example, $^{90}\text{Y}$ (2.21 MeV $\beta$-particles) because of the continuous distribution of detected primary events from unscattered bremsstrahlung photons along the energy scale. Yet, the ESSE method together with properly Monte Carlo simulated scatter kernels have shown to be useful for quantitative bremsstrahlung SPECT imaging (Minarik et al., 2008, 2010).

Recent research on scatter compensation methods that have focused on including Monte Carlo simulation directly into the iterative reconstruction process during the iterative process calculate the scatter component within the energy window using the patient’s own geometry (taken from a registered CT) and a proper model of the camera. This approach is very promising but has a large drawback in that the method is very computer intensive. For $^{99m}\text{Tc}$ applications, however, clinically acceptable

![FIGURE 15.2](image-url) Simulation of a $^{99m}\text{Tc}$ whole-body study. Image (a) shows the case with no scatter and attenuation in the phantom. This is not a realistic case but can serve as a reference image when comparing different scatter and attenuation compensation methods. Image (b) shows the total image, including interactions in the phantom. This image corresponds to a real measurement. Image (c) shows the scatter component in image (b) and image (d) shows the scatter as has been estimated by the triple energy method. Note that this image has more statistical noise due to the narrow energy windows.
processing times can be achieved when using variance reduction methods such as “forced detection,” which in this case means that the explicit simulation of interactions in the collimator and the camera is replaced by a convolution process using a proper collimator response function (Staelens et al., 2007; deJong and Beekman, 2001). In $^{99m}$Tc imaging, this can be justified since 140 keV photons that interacts in the NaI(Tl) crystal most likely will be absorbed at the first interaction site. For other radionuclides, such as $^{131}$I, the septal penetration and the contribution to the image from photons back-scattered behind the crystal is more difficult to model by this approach but investigations have shown it is possible (Liu et al., 2008). If the time required for the calculation is not critical, then a fast Monte Carlo program, such as the SIMIND code (Ljungberg and Strand, 1989), can be used. Dewaraja et al. have shown such an approach for $^{131}$I tumor treatment planning using quantitative SPECT (Dewaraja et al., 2006). Recent development in GPU-based computing may also speed Monte Carlo simulation that might lead to a clinical use of such a scatter compensation method (Kim and Ye, 2011).

### 15.4.3 Estimation of Collimator Response Function

The collimator in SPECT is optimized for a certain photon energy (low-energy, medium–energy, and high-energy collimators), but if there sometimes occur additional high-energy emissions in the decay, such as in $^{131}$I and $^{123}$I, then septal penetration with resulting characteristic star-like artifacts will occur (Dewaraja et al., 2000; Larsson et al., 2006). This penetration deteriorates the image quality (spatial resolution and image contrast) and makes it difficult to quantify the activity (Figure 15.3). Compensation for septal penetration can be made by adding the effect in the forward projector in the iterative reconstruction. For $^{99m}$Tc, where penetration is a low-probability effect, the collimator response is accurately described by a geometric Gaussian-shaped component but when photon energies are used were septal penetration becomes important then it necessary to consider explicit Monte Carlo simulations of the actual photon interactions in the collimator to establish the amount and the distribution of events from the penetrating photons but also for photons scattered in the collimator.

### 15.5 Absorbed Dose Calculations

The absorbed dose can be calculated from a quantitative SPECT image by basically three methods. First, all energy emitted from the decay is located in the voxel. This corresponds to an absorbed fraction, $\phi$, in Equation 15.2, equal to 1. Second, the distribution of energy can be modeled by convolving the SPECT images with an appropriate point-dose kernel that radially describes the distribution of energy emitted from a point source. Third, the radiation transport can be modeled in an explicit way within a computer phantom of the patient using coupled photon–electron Monte Carlo programs. Although point dose kernels are usually derived from very accurate and well-validated programs, such as EGS4 (Nelson et al., 1985), EGSnrc, Geant4 (Allison et al., 2006), and MCNP5 (Forster et al., 2004), the procedure assumes that the kernel is invariant in respect to spatial changes in the distribution. Thus, the method is not accurate in boundaries between different densities and/or tissues. A full Monte Carlo calculation can account for this nonhomogeneity and can also use the patient-specific geometry since today a SPECT/CT combination of images is often available for these types of patients. The approach of assuming local energy absorption can be justified when the radial ranges of the particles is less than that of the voxel size of the SPECT image (Ljungberg and Sjogreen-Gleisner, 2011).

### 15.6 Diagnostic Dosimetry

In diagnostic dosimetry, the aim is to relate the absorbed dose to the risk for late stochastic effects, such as cancer, when exposed to a low dose of ionizing radiation. A population rather than a specific person is therefore of interest. Therefore, more standardized computer phantoms that simulate a “reference man”...
Monte Carlo Techniques in Radiation Therapy

are used, which will simplify the dosimetry since a set of $S$ values can be precalculated and used as representative of the population. $S$ values have been calculated by Monte Carlo methods for several configurations of the phantom (male, female, 15y, 10y, 5y, 1y, newborn, and a pregnant female at 3, 6, and 9 month) for 117 relevant radionuclides. The $S$ values can be found in specific dosimetry codes, such as the Olinda program (Stabin et al., 2005) or at the RADAR web page (Stabin). The cumulated activity still need to be determined but this is done in work preceding the release of a new radiopharmaceutical.

15.7 Therapeutic Dosimetry

When planning for a radionuclide therapy, the treatment is often preceded by a pretherapy planning phase where a small amount of the radiopharmaceutical is administered to the patient to determine the uptake and expected absorbed dose to tumors and organ at risk per unit administered activity. This estimation requires at least three time point measurements. Imaging with SPECT is superior because it is a 3D method and the problem with over- and underlapping activity that is common in planar imaging can be avoided and compensation methods are generally more accurate. The time–activity curves for the VOI (volume of interest) are determined by analyzing the SPECT images using user-defined region of interests. If the SPECT images can be registered to each other, then it is possible to obtain suborgan kinetic curves that show activity variations within the organ. If these are integrated over time to calculate the cumulated activity, an absorbed dose image can be obtained by also applying any of the three absorbed dose calculation methods described above. The absorbed dose in the VOI per unit administered activity will then be the base to calculate the activity for therapy expected to give a prescribed absorbed dose.

Often, the therapy radionuclide decays with photon emission not suitable for SPECT imaging. For example, $^{90}$Y is a useful $\beta$ emitter but the decay lacks a gamma photon emission useful for imaging, which makes this radionuclide not easy

![Energy pulse-height distribution—$^{131}$I](a) ![Energy pulse-height distribution—$^{123}$I](b) ![Energy pulse-height distribution—$^{90}$Y](c) ![Energy pulse-height distribution—$^{99}$Tc](d)

**FIGURE 15.4** Monte Carlo simulated energy pulse-height spectrums for $^{131}$I (a), $^{123}$I (b), $^{90}$Y (c), and $^{99}$Tc (d) using the phantom and distribution shown in Figure 15.2. The dashed line represents events from photon unscattered in the phantom and that are geometrically collimated and thus represents the position of the decay.
to use in a pretreatment planning. Therefore, surrogate radionuclides usually are used. In treatment with $^{90}$Y-labeled rituximab antibodies for non-Hodgkin's lymphoma, $^{111}$In is used as a pretherapy radionuclide. It must here be assumed that the biokinetics is similar between the two compounds. The equivalent absorbed dose for $^{90}$Y is then calculated based on the activity measurement of $^{111}$In. Taking into account the differences in decay scheme and half-life. In recent research, focus is on using the bremsstrahlung photons emitted as part of the slowing down of the $\beta$-particles. However, the acquired bremsstrahlung spectra by SPECT systems are complicated to analyze since no clear photopeaks appear but rather a continuum of scattered and primary events making the selection of energy window locations for SPECT imaging difficult. However, Monte Carlo simulation of the whole imaging process can be still performed with dedicated software that allow for detailed analysis of the components to build an energy spectrum (Figure 15.4) and a corresponding image. This makes it possible to optimize the size and location of the energy window.

Monte Carlo methods can also be used to create kernels that are used for scatter compensation and collimator-response compensation. Minarik et al. (Minarik et al., 2008; Minarik, 2010) have developed a method for activity quantitation based on Monte Carlo simulated kernels and found good accuracy in a phantom experiment comparing activity values estimated from SPECT images with known values.

15.8 Monte Carlo Codes for Simulating Nuclear Medicine Systems

Several Monte Carlo programs have been developed specially for nuclear medicine imaging simulation, and others are public domain programs mainly developed for complex electron–photon simulation but can be adapted for nuclear medicine imaging simulations.

15.8.1 SIMIND Code

The SIMIND Monte Carlo program has been developed by Michael Ljungberg at the Lund University and is a photon program, dedicated to simulate scintillation camera imaging and SPECT (Ljungberg and Strand, 1989; Ljungberg, 1998). It is written in Fortran-90 and is available for Windows, Mac, and Linux operating systems from the web page (Ljungberg, 2011). It simulates photon interactions and is optimized by the use of several variance reduction methods, such as forced detection, photon splitting, delta scattering, and importance sampling. It simulates both voxel-based phantoms and more simple geometrical shapes. Transmission imaging is also possible to simulate. The program has been used for evaluation of scatter and attenuation compensation method, receiver operation characteristic (ROC) analysis, and to evaluate planar activity quantification methods. This code has been used to generate the examples included in this chapter.

15.8.2 SimSET Code

The software simulation system for emission tomography (SimSET), developed at the University of Washington by Robert Harrison et al. (Harrison and Lewellen, 2012), is designed for both single-photon emitters (SPECT and planar systems) and positron emitters (PET systems). The package includes a photon history generator (PHG), the object editor, a collimator module, and collimator, detector and binning modules. The PHG tracks photons through the tomographic FOV creating a list of photon histories. The object editor helps the user to define the activity and attenuation objects for the PHG expressed as voxel images. The collimator module simulates PET collimators and SPECT collimators. The detector module simulates typical PET and SPECT detectors including energy and (for PET) time-of-flight resolution. The binning module allows for binning in terms of number of scatters, axial position, transaxial distance and angle, and photon energy. The SimSET package does implement techniques to improve efficiency. To gain access to the code, and to be added to a list of users for update notices, interested individuals should send electronic mail to simset@u.washington.edu.

15.8.3 Gate Code

The GATE program (Jan et al., 2004) was originally dedicated to simulation of PET and SPECT systems but from version 6, computed tomography, conventional radiotherapy, and hadron therapy can also be simulated (Jan et al., 2011). This open source simulation software is developed by the OpenGATE collaboration (http://www.opengatecollaboration.org). It is based on the Geant4 Monte Carlo tool kit and uses Geant4's libraries to simulate the particle transport. The user employs an extended version of the Geant4 script language that means that no programming is necessary to set up a simulation, unless the user wants to add or modify the features in the GATE program. The program has a layered architecture with a core layer that defines the main tools and features of GATE with routines written in C++, an application layer with C++ base classes, and at the top, a user layer where the simulations are set up using command-based scripts. An original feature is the possibility of simulating time-dependent phenomena such as source kinetics and movements of geometries, for example, patient motion, respiratory and cardiac motions, changes of activity distribution over time, and scanner rotation. Geant4 does however require static geometries during a simulation. GATE has a very powerful geometry package that allows for simulation of very complex geometries such as modern PET camera and prototypes. It also allows for simulation of analytical or voxel-based geometries and output results to various file formats. GATE is currently the only Monte Carlo simulation tool supporting both imaging and radiotherapy applications. A large number of SPECT and PET scanners have already been modeled in GATE, as well as a number of linear accelerators (see http://www.opengatecollaboration.org for a list of modeled systems). This software is used for assisting detector design, assessing acquisition and quantification protocols, and for tomographic reconstruction, in both clinical
and preclinical imaging. GATE also now supports the modeling of CT and optical imaging.

**15.9 Voxel-Based Phantoms**

A voxel-based phantom is described as images with small voxels representing either activity or density (tissue composition). The main advantage with these types of phantoms is the ability to simulate very complicated geometries from a large number of small voxels. The use of voxel-based phantoms in Monte Carlo program requires some form of ray-tracing method, which makes the computing more demanding as compared to analytical phantoms. Ray-tracing can be made by dividing the particle track into equal spacing and during the step process collect anatomical information about the tissue composition and density. Other methods, such as the delta scattering method (Woodcock et al., 1965), have shown to be effective. The drawback with voxel-based phantoms, derived from, for example, a computer tomography study, is that it is relatively difficult to make internal changes of the phantom, that is, adjusting volumes and location of individual organs. Hybrid phantoms are the latest development that combines the advantages of voxelized and mathematical phantoms. Hybrid phantoms are developed from patient data, from either CT or MRI, which are segmented to an initial voxelized model. These segmented structures are then fit with smooth surfaces, such as nonuniform rational b-splines (NURBS) or different types of meshes based on polygons. Based on patient data, the NURBS surfaces or polygon meshes can accurately model each structure in the body, providing the realism of a voxelized model. They also have the flexibility of a mathematical phantom to model anatomical variations and motion. NURBS surfaces can be easily modified through the control points that define their shape while the polygon meshes can be easily altered through the vertex points that define the polygons. These surfaces can then be redefined in a simple manner by adjusting control points instead of each individual voxel. In the final step, the surfaces are used to define new voxel-based images on arbitrary resolution than then can be used as input to a Monte Carlo program. An example of such a hybrid phantom is the 4D NURBS-based XCAT phantom developed by Segars (Segars et al., 2010). Figure 15.5 shows rendered images of the XCAT phantom.

The XCAT phantom actually is a software package. The user modifies an input file with a large set of parameters that control the phantom dimension, organ activities, respiratory movements, and so on. Also specified in that file is the resolution and pixel size for the two sets of images (photon attenuation and activity distribution) that the software package produces. These images can then be used in Monte Carlo programs. It should be remembered that the activity is usually uniformly sampled within an organ defined by the voxel phantom. Also, the attenuation is generally treated as uniform. If more realism is needed, then an organ needs to be further segmented into compartments. A comprehensive textbook on the use of voxel phantoms in medical applications has recently been compiled by Xu and colleagues (Xu and Eckerman, 2009).

Figure 15.6 shows an example of a whole-body scintillation camera simulation using the XCAT phantom and the SIMIND Monte Carlo program. The figure illustrates the degradation in image quality caused by collimator penetration, and scatter and attenuation effects in the phantom and camera compartments, but which is possible to evaluate using detailed Monte Carlo simulations.

**15.10 Summary**

As in many areas in the field of radiation physics, the Monte Carlo method in nuclear medicine has provided a very important insight in how the quality of images depends on the radiation transport and limitation in detector system. The simulation of detector systems in combination with very realistic mathematical phantom is today able to predict experimental measurements to a high degree of accuracy. With increasing development in computing science and hardware development, the method will be very important for “real-time” calculation.
of, for example, scatter models in iterative reconstruction algorithms. A field not mentioned here is the possibility of using Monte Carlo simulated data for educational purpose when expected detectability and recognition of possible false-positives due to physical effects, such as attenuation and scatter, could be compiled as a training package.

References


Minarik, D. 2010. Activity quantification based on scintillation camera imaging—Application to 111In/90Y radioimmunotherapy Department of Medical Radiation Physics, Lund University, Lund.


16.1 Introduction

Since the discovery of x-rays by Rontgen in 1895, x-ray imaging has been widely used in medical applications. The advent of x-ray computed tomography (CT) imaging was one of the most significant advancements in radiography. CT images are obtained by rotating a highly collimated x-ray source around the patient and having detectors record projection x-ray data from a multitude of angles throughout the rotation. Tomographic image reconstruction techniques are then used to create an axial cross-section image of the irradiated volume of the patient. Three-dimensional (3D) images can be constructed by translating the patient through the plane of the x-ray beam in either a stepwise or a continuous motion, resulting in axial or helical scans, respectively, which can be combined in a single 3D image set. Volumetric patient CT images are commonly used in radiotherapy treatment planning. In recent years, a new technique in imaging-guided radiotherapy (IGRT) is kV cone-beam computerized tomography (CBCT) (Uematsu et al., 1996; Jaffray et al., 1999, 2002; Siewerdsen and Jaffray, 2000; Siewerdsen and Jaffray, 2000; Shiu et al., 2003; Yenice et al., 2003; Sharpe et al., 2006) of the patient on the treatment table. CBCT is capable of providing accurate 3D volumetric knowledge about the patient’s anatomy for every treatment fraction and is also suitable for adaptive corrections of errors related to interfractional uncertainties of the treatment process. With the availability of in-room kV-CBCT, it becomes practical to monitor the patient treatment positions and tumor shrinkage on a daily or weekly basis as well as to perform potentially adaptive radiotherapy (ART) (Ding et al., 2007b; Pawlowski et al., 2010) and IGRT (Jaffray et al., 1999; Jaffray and Siewerdsen, 2000).

The utilization of x-ray imaging in IGRT has dramatically improved the radiation treatment and improved the lives of cancer patients. X-ray image guidance has emerged as the new paradigm for patient positioning and target localization in radiotherapy. Although widely varied in modality and method, all radiographic guidance techniques have one thing in common, the use of ionizing radiation. Monte Carlo techniques have been extensively applied in both imaging and radiotherapy physics in recent decades. It is well accepted that Monte Carlo methods offer the most powerful tool for modeling radiation transport for radiotherapy applications (Verhaegen and Seuntjens, 2003). There are several advantages of using Monte Carlo techniques in radiation dosimetry and in the development of new medical devices for treatment delivery and for x-ray imaging. In the following, the application of Monte Carlo techniques will be described.

16.2 Monte Carlo and Kilovoltage X-Ray Imaging

Monte Carlo techniques play an important role in the development of new image devices and radiation dose calculations to patients resulting from an imaging procedure using kilovoltage x-rays.

The knowledge of the characteristics of x-ray beam, including the scattered particles, is important for optimizing a scanner design geometry (Siewerdsen and Jaffray, 2000). Although theoretical scatter models involving both experimental measurements and mathematical algorithms have generally been used in many studies (Tofts and Gore, 1980; Glover, 1982; Johns and Yaffe, 1982; Joseph and Spital, 1982; Merritt and Chenery, 1986;
16.2.1 Simulations of Realistic X-Ray Sources

With many improvements in the Monte Carlo codes, such as in EGSnrc (Kawrakow and Rogers, 2002), it is possible to accurately model low-energy x-ray beams in the diagnostic energy range. In addition, with ever-increasing computing speed and decreasing cost of a computer, it is possible to perform the simulation within an acceptable time. Furthermore, the development of a special-purpose Monte Carlo code, such as BEAM (Rogers et al., 1995), to simulate radiotherapy beams has made it practical to simulate not only radiotherapy beam but also beams from x-ray tube. The BEAM code has been extensively used to simulate the electron and photon beams in commercial medical accelerators (Nelson et al., 1985b; Rogers et al., 1995; Kawrakow and Rogers, 2002). The BEAM code (Rogers et al., 1995) has been used extensively to characterize both megavoltage electron and photon beams as well as kilovoltage x-rays (Ding and Rogers, 1995; Verhaegen et al., 1999; Ding, 2002; Sheikh-Bagheri and Rogers, 2002a,b; Ding et al., 2006, 2007a; Jarry et al., 2006). In the Monte Carlo simulation, one can obtain the information that typically includes data pertaining to the particle’s position, direction, charge, and interactions that it has undergone. In the BEAM code, this information is stored in the phase-space file using a variable, LATCH. The ability to trace a particle’s history allows the fluence of particles that reach the detector panel to be broken down into components: fluence of particles that have interacted only in certain regions, fluence of particles that have not interacted in certain regions, and fluence arising from particles with other user-specified LATCH bit settings. By using LATCH, the Monte Carlo simulation is capable of separating the scattered particles from the primaries even when they have the same incident angle and energy at the surface of the detector. This is the most effective method to obtain details of interactions within the phantom when x-ray passes through a phantom or a patient. By analyzing the simulated results, we cannot only obtain the x-ray properties before and after an x-ray beam interacts with imaged target but also reveal the characteristics of both primary and scattered photons and their dependence on the size of the imaged target.

The characteristics of x-ray beams depend on scan protocols in which both kVp and filters differ; the x-ray source with detailed x-ray tube geometry for different scan protocols has to be modeled in detail.

Figure 16.1 shows examples of the geometry of a Varian OBI system for the two kV-CBCT scan modes where both bow-tie filters and x-ray collimations differ. The Monte Carlo technique is capable of simulating each realistic x-ray source based on a selected parameter of each image acquisition procedure. The Monte Carlo simulation provides realistic beam details as shown in Figures 16.2 and 16.3 in which energy spectra and the mean energy distributions as a function of the central axis for the five different kV-CBCT beams are presented. For each kV-CBCT beam, two curves are presented: one is analyzed photons within the field defined by the X-Y collimator blades openings (within beam field) and the other is within a small 4 × 4 cm² field near the beam central axis (within 4 × 4 cm²). It is seen that the beam spectra near the central axis (where the thickness of the bow-tie filter is thinner) contain more low-energy photons that is consistent with the observation shown in Figure 16.3. The accuracy of each Monte Carlo simulated beam can be validated against measurements as shown in Figure 16.4. It has been shown that the information provided from Monte Carlo simulation plays a significant role in configuring a kV-CBCT beam in a treatment planning system (Alaei et al., 2010).

Low-energy x-rays are often used in imaging guidance techniques. Unfortunately, low-energy x-ray dose calculation methods are not a part of current radiotherapy treatment planning systems and therefore, this additional imaging dose cannot be included in the calculation of total dose to radiotherapy patients. The Monte Carlo simulations are not only capable to accurately calculate radiation doses to patients resulting from an imaging guidance procedure but also provide beam data for configuration of an x-ray beam in a commercial treatment planning systems (Alaei et al., 2010). The information of additional imaging dose to patient’s sensitive organs helps clinicians to make informed decisions on patient selection and the frequency of imaging guidance for each patient. This can lead to improved radiotherapy outcomes and reduced serious complications and poor outcomes of radiation treatment.

The x-ray source parameters obtained from Monte Carlo simulations shown in Figures 16.2 and 16.3 are difficult to obtain through the experimental methods. In addition, the methodology has developed to calculate and predict absolute dose to patients based on CT images from any x-ray imaging procedure (Ding et al., 2008a).
FIGURE 16.1 Schematic drawing to illustrate: (a) Coordinates used in this study, (b) the OBI scan geometry in full fan mode where the detector is centered, and (c) the OBI scan geometry in half fan mode, where the detector is off-set toward the positive X-direction. The isocenter of the CBCT is at 100 cm from the x-ray source and the source-to-detector distance can be selected from among 140, 150, and 170 cm distances. The standard focus to detector distance of 150 cm is employed in our study. The CsI flat panel detector with an active area of $40 \times 30 \text{ cm}^2$ has an anti-scatter grid and the grid lamellas run only along the long axis (X-direction), which is mounted directly above the detector. (Reproduced from Ding G X and Coffey C W 2010. *Phys Med Biol* 55 5231–48. With permission.)

![Schematic drawing](image)

16.2.2 Patient Dose Resulting from Imaging Procedures

The benefits of CBCT imaging are immense because it is possible to see the target and organs at risk before treatment. However, the increased use of x-ray imaging, such as CBCT in radiotherapy, for daily imaging procedures for patient setup, will add significantly to the doses to the normal tissues of the patient (Islam et al., 2006; Gayou et al., 2007; Wen et al., 2007). Accurate dosimetry of radiation dose from image guidance is becoming increasingly important for radiation oncologists to make informed decisions regarding the increased dose to radiosensitive organs. Unlike the general situation with diagnostic imaging, IGRT adds the imaging dose to an already high level of therapeutic radiation. The radiotherapy community is not alone in its search for an appropriate solution for the rapid development of increased utilization of x-ray imaging. The benefits of diagnostic imaging have revolutionized the practice of medicine; however, it has resulted in a significant increase in the population’s cumulative exposure to ionizing radiation. The American College of Radiology (ACR) convened an ACR Blue Ribbon Panel on Radiation Dose in Medicine to address this issue. A recent white paper publication (Amis et al., 2007) has detailed a proposed action plan for diagnostic radiology community. In diagnostic imaging, the primary concern is that the expanding use of imaging modalities using ionizing radiation may eventually result in an increased incidence of cancer in the exposed population. In radiotherapy application of IGRT, the concerns include both the secondary cancer incidence and the increased dose to radiosensitive organ. Hence, the management of imaging dose during radiotherapy will be different from that of routine diagnostic imaging procedures. Recent publication of the AAPM TG-75 report (Murphy et al., 2007) has provided estimated imaging dose data for a variety of imaging guidance techniques and recommended optimization strategies to trade off imaging dose with improvements in treatment delivery. Recent ICRP Publication-102 (ICRP-102, 2007) regarding patient dose in multidetector CT has emphasized that it is important that patient dose is given careful consideration, particularly with repeated or multiple examinations. Experimental methods (Islam et al., 2006; Gayou et al., 2007; Wen et al., 2007) have been used to estimate the radiation dose to patients from CBCT. However, it is difficult to accurately measure dose to different parts of the body, especially when there are large variations in anatomic numbers between bone and soft
The Monte Carlo method has its advantage in providing detailed information on the overall shape of the dose distributions in a real patient anatomy. In the last section, it is demonstrated that the Monte Carlo technique is able to accurately simulate kV x-ray sources. By using the Monte Carlo simulated realistic x-ray source, the dose distributions to a patient can be calculated using CT-based volumetric images.

Although new model-based calculation algorithms and methodologies are being developed for radiotherapy treatment planning system to model the radiation dose from low-energy x-ray beams (Ding et al., 2008c; Pawlowski and Ding, 2011), the Monte Carlo technique is the most accurate calculation method in the dose calculations for kV beams. Because of the presence of the higher-Z materials in the body, such as bone, the current available model-based dose calculation algorithm in commercial treatment planning systems is adequate for patient dose calculations due to the increased photoelectric effect at low x-ray energies. Figure 16.5 illustrates the photoelectric effect on the calculated dose distributions for a patient treated for head and neck cancer. The isodose lines are normalized in such a way that the maximum dose is 50%. It is seen that the isodose lines above 20% are in the regions of bony anatomy. The dose to the bone is approximately 3–4 times the dose to soft tissues. It is evident that density correction alone is not adequate for dose calculations for x-ray in diagnostic energy range. Before the new accurate model-based dose calculation algorithms are developed, Monte Carlo techniques remain to be the most effective method available.

To accurately calculate the absolute imaging dose for a patient resulting from an image acquisition procedure, the simulated x-ray source has to be calibrated. The calibration of simulated source is necessary for absolute dose calculation to patients. The calibration algorithm allows the user to calculate both relative and absolute radiation doses to patients. In the next section, a method to calibrate simulated beams will be described.

16.2.3 Calibration of X-Ray Beam Output

To accurately calculate absolute dose received by a real patient during a CBCT image acquisition, a technique was developed to calibrate Monte Carlo calculated dose (Ding et al., 2008a). It is accomplished in two steps. First, the absolute dose to a
point in a phantom with known geometry is measured by using an ionization chamber based on the following equation (Ma et al., 2001):

\[ D_w = MN_i P_{Q,cham} \left( \frac{\bar{\mu}_m}{\rho} \right)^w_{air} \]  

(16.1)

where \( D_w \) is the dose-to-water at the point of measurement in a water phantom, \( M \) is the leakage-corrected chamber reading in coulombs (C), \( N_i \) is the air kerma calibration factor for the given beam quality, \( Q \), in Gy/C, \( P_{Q,cham} \) is the overall chamber correction factor that accounts for the change in the chamber response due to the displacement of water by the ionization chamber and the presence of the chamber stem, the change in the energy, and so on, and \( \left( \frac{\bar{\mu}_m}{\rho} \right)^w_{air} \) is the water-to-air ratio of the mean mass energy-absorption coefficients. The air kerma calibration factor is traceable to national standards, that is, from an Accredited Dosimetry Calibration Laboratory (ADCL), National Institute for Standards and Technology (NIST), or National Research Council of Canada (NRCC). In this study, air kerma calibration factors for the specified kV-CBCT beam are obtained from an ADCL.

The simulated realistic CBCT beam is stored in a phase-space file (Rogers et al., 1995). The calibration of a Monte Carlo simulated beam involves two steps. First, using the simulated beam, the dose to a point in the phantom from a CBCT scan is calculated to be \( D_{MC,cal} \), which is in units of Gy per incident particle (on the x-ray tube). Using an ionization chamber, the dose to the same point in the phantom is experimentally determined based on the Equation 16.1 to be, \( D_{exp} \), in Gy for the same CBCT acquisition. Second, we introduce a Monte Carlo calibration factor, \( f_{MC,cal} \) in the following equation:

\[ D_{exp} = f_{MC,cal} D_{MC,cal} \]  

(16.2)

The factor, \( f_{MC,cal} \) is unique to each Monte Carlo simulated beam stored as a phase-space file. A simulated x-ray beam is called “calibrated” when the beam’s calibration factor, \( f_{MC,cal} \) has been determined through Equation 16.2. Using the calibrated CBCT beam, the Monte Carlo calculations can be used to determine the absolute dose, \( D_{MC} \), to any point in a different phantom from the same simulated beam using the following equation:

\[ D_{MC} = f_{MC,cal} D_{MC,cal} \]  

(16.3)

FIGURE 16.5 (See color insert.) (a) Monte Carlo calculated dose distributions (isodose lines) on CBCT image slices for a patient treated for head and neck cancer. (b) Dose profiles along lines A, B, C, D, E, and V. Note that line A crosses the eyes and the bony part of the nose, and B crosses the brain. Line C crosses the mandible. Lines D, E, and V cross the spinal cord at different locations. (Reproduced from Ding G X, Duggan D M, and Coffey C W 2008a. Med Phys 35 1135–44. With permission.)
where \( D_{MC} \) is the Monte Carlo predicted dose in Gy. The method developed here is not limited to kV-CBCT beam and can be applied to megavoltage photon and electron beams as well. However, the factor \( f_{MCcal} \) is specific to the control quantity. In this case, it is controlled by the specific CBCT acquisition parameters, for example, kV beam’s mA, bow-tie filters, and so on. When these acquisition parameters remain the same, Equation 16.2 can be used to predict the dose accurately for imaged target that is different from the calibration phantom.

16.2.4 Method to Determine an X-Ray Source Output for Imaging Dose Calculations in Treatment Planning

In addition to kV-CBCT beam parameters (such as dose profile, beam spectrum, etc.), the determination of the beam output is required to commission a radiation beam in a commercial treatment planning system. This is similar to the method used in the last section to calculate the dose to the patient resulting from a kV-CBCT procedure. In a commercial treatment planning system, the beam outputs for a specific kV-CBCT scan determined by measured values in phantom were entered in the system as cGy/min during the beam configuration (Alaei et al., 2010), with 1 min being equivalent to a single kV-CBCT scan. This type of output configuration is treatment planning system specific.

It is worth noting that there is a conceptual difference between the radiation dose calculation to a patient using a therapeutic beam and a kV-CBCT beam in a treatment planning system. The dose calculation from a therapeutic beam is to determine the monitor units, and the shape and the intensity of beams to cover the prescribed radiation dose to the targeted tumor volume, whereas the dose calculation from a kV-CBCT beam is to determine the dose to a patient resulting from a specified kV-CBCT scan. For a given kV-CBCT scan, the dose differences between different patients (or between different scanned sites of the same patient) result from the geometry differences of the scanned patient body; the kV-CBCT beam is unchanged.

The purpose of the measurement and the calculation of the reference dose resulting from a kV-CBCT scan is to determine “the kV-beam output” for a certain reference situation in a treatment planning system so that the calculated dose at the point of measurement equals the measured dose to the reference phantom.

Unlike MV photon beams where dominant interactions are Compton effects, the dose from kV beams is sensitive to phantom atomic number \( Z \) due to the increased photoelectric effect. Therefore, the water phantom will be the natural choice for the reference dose measurement; however, plastic phantoms are more practical and convenient in experimental setups. Since Monte Carlo calculations are generally not available for photon beams in commercial treatment planning systems, it is important
to find the error in the dose calculations introduced without considering medium dependence. In a commercial treatment planning system that uses model-based dose calculation algorithms (Mackie et al., 1985; Mohan and Chui, 1987; Ahnesjo and Aspradakis, 1999), the dose calculations are based on electron density, which is converted from CT number. Therefore, the accuracy of dose calculations without considering the medium dependency affects the accuracy of beam output determination of the kV-CBCT beams. It has been demonstrated that in Monte Carlo simulations the composition of medium must be taken into account and it is important to select a suitable water substitute material in order to reduce the error in the absorbed dose determination (Ding and Coffey, 2010). This was done by comparing calculated doses at the point of measurement in a plastic phantom between Type I and Type II calculations.

Several investigations have demonstrated that the Monte Carlo simulation could provide details of an x-ray beam data that were essential for commissioning an x-ray beam in a radiotherapy treatment planning system (Alaei et al., 2010). The Monte Carlo generated beam parameters and method of beam output calibration make it possible to commission a kV-CBCT beam in a radiotherapy treatment planning system (Alaei et al., 2010), which facilitates reporting and accounting the additional radiation doses to radiotherapy patients resulting from repeated kV-CBCT procedures (Ding and Coffey, 2008, 2009).

16.2.5 Applications

16.2.5.1 Developing and Validating New Model-Based Dose Calculation Algorithms

The increasingly intensive imaging procedures for IGRT now obligates the clinician to evaluate the total dose to the target and radiosensitive organs from both therapeutic and imaging doses to avoid significant underestimation of dose to radiosensitive organs. This requires that radiotherapy treatment planning systems be capable of calculating dose from kilovoltage photon beams, since low-energy x-rays are often used in imaging guidance procedures. Unfortunately, low-energy x-ray dose calculation methods are not a part of current radiotherapy treatment planning systems and therefore, this additional imaging dose cannot be included in the calculation of total dose to radiotherapy patients. Attempts have been made to incorporate the imaging dose into total patient dose from diagnostic x-rays in treatment planning by Alaei et al. (2000). Owing to the increased photoelectric effect, a kV-CBCT procedure results in doses to bone that are up to a factor of 3–4 higher than those in surrounding soft tissue (Verhaegen et al., 2004; Ding et al., 2008b). It has been shown that the model-based (e.g., convolution/superposition) algorithm is capable of computing dose within phantoms of density equal to or less than that of water and that its limitations introduce inaccuracies for calculations through moderately high-atomic-number materials such as bone for kV x-rays (Alaei et al., 1999, 2000, 2001; Seuntjens and Ma, 1999; Verhaegen et al., 2004). Currently, the Monte Carlo technique is the most accurate dose calculation method for x-rays in the diagnostic energy range in media where there are large atomic number variations, such as a patient body. However, the lack of availability of the Monte Carlo method in commercial radiation treatment planning (RTP) systems and its long computation time limits its usage for routine patient organ dose calculations. Hence, there is a need to develop a more accurate dose calculation method that is suitable for the implementation in treatment planning systems. Monte Carlo techniques can serve as gold standards in developing and validating new model-based dose calculation algorithms. The progress has been made to develop model-based methods to overcome the deficiencies of existing algorithms in kilovoltage dose calculations, especially for CT-based human volumetric images used in RTP (Ding et al., 2008c; Pawlowski and Ding, 2011).

16.2.5.2 Improving the Design of New X-Ray Tube or Detector to Reduce the Imaging Dose to Patients

Monte Carlo techniques have been shown to be a powerful tool to analyze the beam characteristics and to assist the development of radiation treatment machines (Rogers et al., 1995) as well as x-ray units used in the IGRT (Ding et al., 2007a; Spezi et al., 2009; Ding and Coffey, 2010; Ding and Munro, 2011). It has been shown that the Monte Carlo calculations can be used in modifying and selecting image acquisition parameters to minimize radiation exposures to the patient’s sensitive organs (Ding et al., 2010). In addition, Monte Carlo techniques are capable of calculating the dose to organs such bone marrow (Walters et al., 2009) in which the experimental methods are difficult or impossible for low-energy x-rays.

16.3 Monte Carlo and MV Imaging

16.3.1 General Introduction

Monte Carlo techniques have a long history of being used to model the detection of ionizing radiation. Many of the Monte Carlo codes that are applied to medical physics problems were developed at large nuclear and high-energy particle physics laboratories such as CERN, Los Alamos, and the SLAC National Accelerator Laboratory (Menlo Park, CA) (Nelson et al., 1985a; Solberg et al., 2001; Agostinelli et al., 2003). The use of Monte Carlo in this context is useful for two main purposes: First, in the design and development stage, the radiation detector can be cost-effectively optimized for a particular application. Second, Monte Carlo can be used to predict the response of a radiation detector in a particular experiment for direct comparison with experimental data (Apostolakis et al., 2009; Dunn and Shultz, 2009). This capability is also of use in medical physics applications, including diagnostic and radiation therapy imaging (Rogers, 2006). Verification of radiotherapy treatments is extremely important for ensuring the accurate and precise delivery of the prescribed dose to the correct volume and Monte Carlo techniques have had a role to play in this. In the following sections, some of the ways in which Monte Carlo can be used to verify the different aspects of a radiotherapy treatment will be briefly described. The focus will be on the use of
Monte Carlo modeling of megavoltage electronic portal imaging devices (EPIDs) for verifying patient position and dosimetry.

16.3.2 Technology of Electronic Portal Imaging

EPIDs were developed as a digital replacement for radiographic film to assist in minimizing the patient setup error in a radiotherapy treatment. The detectors are attached to the linear accelerator gantry and can be moved to a position behind the patient for imaging the patient when required. The imaging can be performed immediately before the start of a treatment fraction or even during a treatment fraction. The image of the patient in the treatment position can then be compared with a reference image of the patient position. The reference image was traditionally a simulator radiographic film but is now more likely to be a digitally constructed radiograph calculated using the treatment planning CT. Three distinctly different EPID systems have been developed for commercial use over the last 20 years; the scanning liquid ion chamber array, camera-based systems, and more recently amorphous silicon flat panel detectors (Antonuk, 2002; Langmack, 2001; Kirby and Glendinning, 2006).

16.3.2.1 Scanning Liquid Ion Chamber Arrays

The first widely available commercial EPID was the scanning liquid ionization detector (Meertens et al., 1990). The detector consists of two planes of 256 electrodes separated by a gap of 0.8 mm. The gap between the electrode planes is filled with a liquid, 2,2,4-trimethylpentane, that is ionized when irradiated. The electrode spacing is 1.27 mm giving an active area of 32.5 × 32.5 cm². One plane of electrodes is connected to 256 independent electrometers for reading the collected ionization charge and the other plane of 256 electrodes are connected to a 500 V supply that can apply a high voltage to each electrode individually. During irradiation, the voltage is applied to each electrode in succession (on the voltage plane) with an integration time of around 5 ms and the signal recorded for each of the electrodes giving a total readout time of 1.25 s. The scanning readout is synchronized to the linac pulse signal to avoid introducing artifacts due to any changes in the dose rate during acquisition. The system is compact and is geometrically accurate. The main disadvantage of the system is that during acquisition only a single electrode has a high voltage applied at any one time and as a result is relatively inefficient. The slow ion recombination rate of ~0.5 s compensates for this somewhat but a relatively high patient image dose is still required to obtain a useful image. The detector also has relatively poor spatial resolution. Spezi showed that full Monte Carlo simulations of the linear accelerator and detector response were possible and that measured and simulated dose images agreed to within 2%/2 mm (Spezi and Lewis, 2002).

16.3.2.2 Camera-Based Systems

Several different camera-based systems were developed to increase the active detector area, and improve the sensitivity and spatial and contrast resolution (Althof et al., 1996; de Boer et al., 2000). The camera-based systems incorporate an x-ray to optical light converter; typically, a metal plate and a phosphor screen or solid-state scintillator crystal; a mirror and a camera and lens then capture the optical image. These systems are generally quite large and bulky and have to be physically removed from the linac gantry when not in use. The camera-based systems are also not particularly efficient at collecting the light produced by the phosphor screen and reflected into the camera lens system by the mirror, typically only 0.1–0.01%. This poor light collection efficiency introduces noise into the image. Light collection efficiency can be improved by increasing the thickness of the phosphor screen, thereby increasing the amount of generated light or increasing the solid angle cone of light collected using a large aperture lens and digital camera with a large-area CCD chip (Mosleh-Shirazi et al., 1998). Monte Carlo studies have been performed of the screen–metal plate combination with the aim of optimizing the detector for imaging with a radiotherapy beam produced from a low-Z aluminum target (Flampouri et al., 2002). Yeboah and Pistorius investigated the sensitivity of the detector response to different metal plate–screen combinations and changes in the exit beam spectra from a phantom (Yeboah and Pistorius, 2000). The response of the detector was shown to depend on a combination of phantom–patient thickness, air gap between the detector and phantom–patient, and thickness of the metal plate–screen.

16.3.2.3 Flat Panel Amorphous Silicon Detectors

Owing to the limitations of the liquid ionization chamber and camera-based systems, flat panel amorphous silicon (a-Si:H)-based detectors are now the system of choice on most modern linear accelerators for MV imaging with the first commercial systems becoming available in 2000 (Antonuk et al., 1998). They combine a foldaway compact design with improved sensitivity, and contrast and spatial resolution, and allow high-quality patient images to be acquired with only a few CGy.

The first two components are similar to the camera-based systems. High-energy x-rays interact in a metal plate, usually 1 mm of copper, producing electrons and lower-energy x-rays. The electrons and x-rays then interact in the ~133 mg cm⁻² phosphor screen, typically gadolinium oxysulfide doped with terbium (Gd₂O₂S:Tb), generating optical photons, some of which escape the exit plane of the screen. Behind the screen, each pixel in the detector array consists of a photodiode on a thin amorphous silicon layer that converts each light photon to an electron–hole pair that is stored in a capacitor element.

Each pixel is connected to a thin film transistor switch that is used to control the readout of the stored charge signal. The signal is read out by applying a bias voltage to the transistor, making it conducting. Data are usually read out row by row with each row comprising a data line. The current commercially available systems consist of an array of either 1024 × 1024 or 1024 × 768 pixels corresponding to an area of 41 × 41 cm² or 40 × 30 cm², respectively. The two most common clinical a-Si EPID’s are the Elekta iView GT and the Varian Portal Vision as 500/as1000. The a-Si:H detectors have far superior image quality for a low patient imaging dose compared to the liquid scintillator and camera-based systems.
systems. This is mainly due to the high efficiency of the light collection into the photodiodes. The collected charge signal in each pixel $I_{\text{meas}}(x,y)$ is amplified and digitized. The signal is first corrected for dark current $I_{\text{dark}}(x,y)$ and then for pixel sensitivity or gain using a flood field $I_{\text{flood}}(x,y)$. The corrected signal is

$$I_{\text{corr}}(x,y) = \frac{I_{\text{meas}}(x,y) - I_{\text{dark}}(x,y)}{I_{\text{flood}}(x,y) - I_{\text{dark}}(x,y)}$$

(16.4)

This correction also removes position-dependent variations in the image due to the beam profile from the measured image, and although this will improve anatomical image quality, it is undesirable for accurate dosimetry. The detectors have also been shown to have an overresponse to low-energy x-rays (compared to water) due to an increased photoelectric effect in the high-atomic-number phosphor screen. This is shown in Figure 16.6. This complicates the conversion of the measured signal into dose-to-water for dosimetry applications where Monte Carlo studies have shown that there is significant position-dependent variation in the beam spectrum (Parent et al., 2004). However, the a-Si detector has many properties that make it well suited to dosimetric verification, including a linear dose response (McCurdy et al., 2001; Greer and Popescu, 2003). Two general approaches have been used for using EPIDs for dosimetric verification. The first is known as transit dosimetry where the energy fluence exiting the patient or phantom is determined from a measured image. The energy fluence may then be back-projected through a CT dataset of the patient or phantom and converted to dose before comparing with the planned dose distribution (McNutt et al., 1996a; Partridge et al., 2002; Ansbacher, 2006; Wendling et al., 2009). The second approach involves predicting a portal dose image and comparing it with the measured portal dose image (McCurdy et al., 2001; Siebers et al., 2004; Van Esch et al., 2004; Kairn et al., 2008; Chytik and McCurdy, 2009). Methods for predicting the portal dose image include a full Monte Carlo method that includes the radiation transport through the head of the linear accelerator and the subsequent deposition of dose in a model of the detector (Spezi and Lewis, 2002; Siebers et al., 2004; Parent et al., 2006b). These methods suffer from long computation times, making them difficult to implement clinically. Another approach is to use precalculated dose kernels that are convolved with the fluence incident on the detector. The dose kernels can be determined from Monte Carlo simulations (McNutt et al., 1996b; McCurdy and Pistorius, 2000) or alternatively analytical kernels can be used (Greer et al., 2007).

16.3.3 Monte Carlo Modeling and MV Imaging

In this section, the use of the Monte Carlo technique for modeling MV imaging devices and the physical processes involved in MV imaging will be discussed. The focus will be on a-Si:H detectors as these are now the most prevalent in clinical use.

16.3.3.1 Modeling Electronic Portal Imaging Devices

The first stage in modeling an EPID is building the model in the appropriate Monte Carlo code. This requires the specifications of the actual device. These details can be obtained from the manufacturer under a nondisclosure agreement. The model is generally quite simple, consisting of a number of layers or “slabs” of different materials with each layer having a thickness and composition specified by the manufacturer. A model of the slab geometry Elekta a-Si:H detector can be built using the DOSXYZnrc user code, part of the EGSnrc code. This will be used as an example but the same principles can be used to construct a model for any of the currently available commercial a-Si:H EPIDs. DOSXYZnrc has been developed to calculate dose in a rectilinear voxel-based phantom and a number of authors have applied it to the modeling of a flat panel detector (Spezi and Lewis, 2002; Siebers et al., 2004; Parent et al., 2006b). The active area of the detector is $41 \times 41$ cm$^2$. The actual detector has 1024 x 1024 pixels but a lower-resolution model can be used to
speed up the calculation. An array of 256 × 256 pixels can be used in the model, giving a model pixel pitch of 1.6 mm (this corresponds to a projected pixel size of ~1 mm at the isocenter plane). The voxel thickness in the z direction will be different for each layer and will depend on the thickness of each layer. The number of voxels in the z direction will correspond to the number of layers in the detector. The calculated portal dose image is taken to be the dose scored in the \( \text{Gd}_2\text{O}_2\text{S}:\text{Tb} \) layer of the EPID model. Munro and Bouius showed that for an a-Si EPID, made only of a 0.1-cm Cu buildup plate, a phosphor screen, and the a-Si components (diodes and TFTs) on a glass substrate with no casing or backscatter material, the signal detected by the light sensor is almost entirely due to photon and electron interactions within the copper and the phosphor layers (Munro and Bouius, 1998). Furthermore, Antonuk et al. showed that the response of the light sensor is proportional to the energy deposition in the phosphor (Antonuk et al., 1998). The geometry of the detector model with respect to the linear accelerator is shown in Figure 16.8.

The Monte Carlo model of the EPID should be commissioned to validate that it is an accurate model of the actual detector. This is done by comparing measured EPID data with data simulated using the EPID model to ensure good agreement. The following is a list of suggested data for commissioning an EPID Monte Carlo model:

- Comparison of measured and simulated in-plane and cross-plane profiles for a flood field (26 × 26 cm²).
- Comparison of measured and simulated output factors for a range of different regular field sizes (1 × 1, 2 × 2, 5 × 5, 10 × 10, 15 × 15, 20 × 20, and 26 × 26 cm²).
To ensure the model EPID accurately models the response to radiation scattered from the patient, measured and simulated profiles with different thickness of slab phantom material (e.g., PMMA or solid water) should be compared.

Finally, comparisons of more complex fields with and without phantom material in the field should be performed, for example, intensity-modulated radiotherapy fields.

Correct normalization of the measured and simulated EPID images is important. A relatively simple approach is to perform a cross calibration:

\[
\frac{I_{MC}(x, y)}{I_{MC}^{10 \times 10}} = \frac{I_{meas}(x, y)}{I_{meas}^{10 \times 10}}
\]

(16.5)

where \(I_{MC}^{10 \times 10}\) and \(I_{meas}^{10 \times 10}\) represent the mean pixel values in a 1 x 1 cm² central region of simulated and measured EID images for a 10 x 10 cm² field, respectively. The measured and simulated EPID response can then be compared directly. For simulations of Varian accelerators, a correction is also required for backscatter from the collimators into the monitor chambers (Liu et al., 2000).

16.3.3.2 Pixel Sensitivity Variations in the Measured Image

There is an inherent difference between the measured EPID image and the simulated EPID image, which must be considered before an accurate comparison can be performed. The measured images are flood field corrected to remove the pixel sensitivity variations of the detector; this procedure also removes the beam profile due to the flattening filter. There are two ways of correcting for this difference. One is to introduce a pixel sensitivity map into the simulated images and a second is to subtract only the pixel sensitivity from the measured image. Both methods require separation of the pixel sensitivity from the beam profile. Parent et al. (2006a) compared three previously used methods with a new method for doing this; the methods were

- Use of the standard flood field image
- Use of a flood field image obtained with a slab of plastic water in the beam to remove the beam profile
- Use of a series of small overlapping 10 x 10 cm² fields (considered to be uniform) to construct a pixel sensitivity map
- Use of an analytical function fitted to the Monte Carlo beam profile

The work showed that the above methods gave similar results but were all superior to using the standard flood field image. Greer previously showed that multiple overlapping fields could be used to derive the pixel sensitivity map for amorphous silicon detectors (Greer, 2005). The method involves delivering a radiation field (10 x 10 cm²) assumed to be uniform to the whole detector using a series of translations of the detector. Each pixel in the detector receives a uniform dose and variations in the pixel values can be used to derive a pixel sensitivity map \(I_{pixel}(x,y)\). Dividing subsequent images \(I(x,y)\) by this pixel sensitivity map generates an image that is corrected for pixel sensitivity variations \(I_{corr}(x,y)\).

\[
I_{corr}(x,y) = \frac{I(x,y)}{I_{pixel}(x,y)}
\]

(16.6)

The method of using a thickness of solid or plastic water above the detector to flatten the beam was used by Siebers et al. (2004) to calibrate the Varian as500 detector. The thickness of plastic water to remove the beam profile can be determined using Monte Carlo simulation. The resulting image is a map of the detector pixel sensitivity variation that can be used in Equation 16.4.

16.3.3.3 Modeling Off-Axis Response

The methods discussed previously for removing the pixel sensitivity from the measured EPID images allow direct comparison of dose to the detector with simulated EPID images. In radiotherapy, the dosimetric quantity of interest is dose-to-water. The high-atomic-number components in a-Si flat panel imagers cause a response that is nonwater equivalent. They have been shown to overrespond (compared to water) to low-energy photons due to an increase in x-ray photoelectric effect interactions (Kirkby and Sloboda, 2005). Since the clinical radiotherapy beam has a spatially varying energy spectrum, the response of the detector will also vary spatially. This is a particular problem for treatment techniques involving multiple small irregular off-axis fields such as IMRT or VMAT where the energy spectra can vary significantly from the calibration beam; Figure 16.9 shows the example of the variation in the energy spectra for a 10 x 10 cm² field on both the central axis and off-axis. The off-axis field energy spectrum shows an increase in the number of photons/energy.
low-energy photons below 1 MeV. Parent et al. (2007) showed that this can result in variations in response of up to 29% if the same IMRT field is delivered on- and off-axis. The spatially varying response can be accounted for by measurement of the beam profile using ion chambers or film (Van Esch et al., 2004; Greer, 2005; Wendling et al., 2006) and calculation-based methods (Li et al., 2006; Winkler et al., 2007). The problem with using ion chambers or film to measure the beam profile is that they both have a different response to the EPID and therefore can introduce errors.

16.3.3.4 Modeling Patient Scatter
Scattering of radiation from the patient into the EPID has implications for both anatomical imaging and dosimetry. Scatter will degrade the quality of the image, reducing contrast and spatial resolution. Dosimetry applications generally require the primary dose or fluence in the EPID and scatter will increase the measured signal. A number of authors have used Monte Carlo simulations to investigate the effect of scattered radiation from the patient into the EPID. Swindell and Evans developed a physical model for the scatter-to-primary ratio (SPR) in portal images that agreed well with Monte Carlo calculations for a range of field sizes, phantom thicknesses, and phantom-detector distances (Swindell and Evans, 1996). Jaffray et al. (1994) investigated how the scatter and primary photon fluence is influenced by the scatter geometry and energy spectrum of the incident beam. For a particular geometry, they showed that for a quantum noise-limited detector system (an a-Si EPID can be considered quantum noise limited), the differential signal-to-noise ratio was decreased by 10–20% due to the scatter fraction. Spies and Bortfeld carried out a study of the effect of scattered radiation on megavoltage images for different field sizes and phantom thicknesses (Spies and Bortfeld, 2001). Using Monte Carlo simulations, they developed analytical models that modeled the scatter as originating from a point source at a position within the phantom that varied as a function of field size. Spies et al. (2001) also demonstrated the use of scatter kernels calculated using Monte Carlo to correct for scattered radiation in megavoltage cone-beam CT reconstructions. The kernels take account of both the spectrum of the incident beam and the dimensions of the phantom being scanned. The scatter corrections reduced the effect of cupping artifacts in the reconstructed data and reduced deviations in reconstructed electron densities from 30% to 8%.

16.3.3.5 Modeling Detector Backscatter
A number of authors have shown the Varian a-Si EPID to have a nonuniform response due to the support arm below the active layers in the detectors. Modeling the detector as a homogeneous slab of material will therefore introduce errors. Ko et al. (2004) carried out a Monte Carlo study to investigate the effect of placing water, lead, or copper slabs behind the active layers to reduce the nonuniform response due to the support structure. An experimental study was also performed by the same group to verify the findings of the Monte Carlo and confirmed that 5 mm of lead reduced the nonuniform backscatter to less than 0.5%. There were also improvements in contrast and spatial resolution (Moore and Siebers, 2005). Cufflin et al. (2010) showed how the nonuniform response could be modeled by including a nonuniform structure of water-equivalent material behind the active layers of the EPID model.

16.3.4 Applications
There are two main applications of Monte Carlo for portal dosimetry. The first is pretreatment verification of IMRT treatments and the second is the verification of the delivered dose during a patient treatment, known as in vivo or transmission dosimetry.

16.3.4.1 Pretreatment IMRT Verification
IMRT patient treatment plans are typically verified by measurement before treatment can begin (Ezzell et al., 2003; Galvin et al., 2004). This can be a time-consuming process and usually requires delivering the treatment fields to a suitable phantom so that ion chamber or film dosimetry measurements can be compared with the treatment plan dosimetry. The two-dimensional (2D) fluence for each field can also be measured using either 2D ion chamber or diode arrays that can then be compared to the planned fluence using a gamma analysis (Low et al., 2011). The commercially available 2D dosimetry devices have relatively poor spatial resolution that makes accurate verification of intensity-modulated fields difficult. EPID devices can be used as 2D high-resolution dosimeters that are conveniently available on most modern linear accelerators, making them suitable for IMRT pretreatment fluence verification. A full Monte Carlo simulation of the linear accelerator output and EPID can be used to predict the 2D dose in the EPID, obviating the need to account for the non-water-equivalent energy response of the detector. The predicted response of the EPID can then be compared directly with the measured response. Despite the relatively long calculation times, this has been shown to be an accurate method for validating the delivered fluence for an IMRT treatment (Spezi and Lewis, 2002; Siebers et al., 2004; Parent et al., 2006b; Cufflin et al., 2010). Figure 16.10 shows a comparison of two different measurement techniques of 2D fluence for an IMRT treatment field, a 2D diode array (a) and EPID (b), and two different calculation techniques, treatment planning system (c) and Monte Carlo (d). The circled region in the 2D diode array image highlights a difference where the limited resolution of the detector is unable to detect the small high-dose gradient region. The pixel sizes of the EPID and Monte Carlo model EPID are both 1 mm (at the isocenter plane). A gamma analysis of the EPID measurements against the Monte Carlo calculated fluence images for all seven fields in this particular prostate IMRT treatment had greater than 96% of pixels passing with acceptance criteria of 3% and 3 mm.

16.3.4.2 In Vivo Patient Dosimetry
The methodology of using Monte Carlo-based EPID dose prediction for pretreatment verification can be extended to...
transmission dosimetry. A model of the patient or phantom can be included in the simulation. Figure 16.11 illustrates schematically how the different components of the treatment can be simulated. First, the linear accelerator model is used to produce phase-space file(s) of the treatment field. Second, the phase-space file(s) are then used as input to model the energy deposition in the patient and phase-space file(s) produced at the exit plane of the patient. Third, this phase space is then used to calculate the energy deposition in a model EPID. The standard distribution of DOSXYZnrc does not generate an output phase-space file, so a modification of the code is necessary (Spezi et al., 2001). An alternative that avoids writing out the intermediate phase-space file is to combine the CT-based model of the patient and the DOSXYZnrc model of the EPID; the dashed box in Figure 16.11 shows the components to be combined. This also requires some scripting or coding effort to combine the two DOSXYZnrc phantoms, particularly for treatment fields with nonzero gantry angles where the two phantoms with rectilinear voxel geometries have to be rotated with respect to one another (Chin et al., 2003; Kairn et al., 2011).

References


Monte Carlo for Kilovoltage and Megavoltage Imaging


17.1 Introduction

The favorable physical selectivity of ion beams offers superior tumor-dose conformality with better sparing of surrounding critical organs and healthy tissue with respect to external radiotherapy modalities based on conventional electromagnetic radiation. However, these ballistic advantages come at the expense of increased sensitivity to conventional sources of treatment uncertainties during the fractionated course of radiation therapy, such as patient positioning errors or other geometrical/anatomical variations (e.g., weight loss) in the treatment situation with respect to the planning one. Moreover, the validity of the analytical pencil-beam algorithms typically employed for ion therapy treatment planning and, in particular, the accuracy of the semiempirical calibration curve converting the patient computed tomography (CT) data into water-equivalent ion range (Schaffner and Pedroni, 1998) can only be carefully assessed via measurements in tissue-equivalent substitutes but cannot be directly verified in vivo. Therefore, to account for all the above-mentioned random or systematic sources of errors, cautious safety margins are still used in clinical routine of ion therapy treatment planning. Moreover, when defining the beam directions, it is a common practice to avoid placement of the distal dose fall-off in front of critical organs. Instead, it is preferred to use the less sharp but more reliable lateral penumbra of the ion beam, often sacrificing the utmost achievable dose conformality for the sake of safety. Therefore, tools for the in vivo validation of the ion range and the actually delivered treatment during the fractionated course of radiation therapy would be highly beneficial and might promote full clinical exploitation of the improved ballistic selectivity offered by ion beams for highly conformal tumor therapy.

Nowadays, positron emission tomography (PET) is the only technically feasible method fulfilling the requirement for a three-dimensional (3D), noninvasive, in vivo monitoring of the delivered ion treatment and, in particular, of the beam range in the patient during or shortly after irradiation (Enghardt et al., 2004a). The unconventional application of this nuclear imaging technique to ion beam therapy is based on the detection of the transient pattern of $\beta^+$-activation, which is produced in fragmentation reactions between the incoming ions and the irradiated tissue. Imaging can be performed during beam delivery (“in-beam”) or minutes afterward (“in-room” or “offline”) depending on the availability of the dedicated or commercial PET installation (Shakirin et al., 2011). Owing to the different underlying physical processes, the irradiation-induced $\beta^+$-activity originating from nuclear interactions can only represent a surrogate signal correlated in a not straightforward way to the electromagnetically deposited dose. Therefore, treatment verification is based on the comparison of the measured PET images with a corresponding expectation based on the planned treatment and the fraction-specific time course of irradiation and imaging (Enghardt et al., 2004a; Parodi et al., 2007c). This comparison may enable detection of unpredictable discrepancies between prescribed and actual irradiation, thus offering an independent verification of the entire therapy chain from treatment planning to delivery as well as opening the possibility for intervention prior to the application of the next therapeutic session (Enghardt et al., 2004b). Therefore, PET-based treatment verification can potentially enable safe reduction of the planning margins and exploitation of the distal ion dose fall-off for more conformal treatments with related dose escalation. Moreover, it might offer an important
tool for assuring the correct treatment delivery in the increasingly considered application of ion beams to high-dose hypofractionated therapy (Ertmer and Tsujii, 2007; Matsufuji et al., 2007), where less or even no subsequent fractions are available for compensation of errors. Finally, it can improve confidence of the successful implementation of special beam delivery strategies like gating or tracking in the challenging presence of organ motion, as supported by first encouraging phantom experiments (Parodi et al., 2009).

The calculation of the expected pattern of activation plays a central role in the clinical implementation of PET-based treatment verification. Although promising analytical approaches are being proposed and evaluated by several groups for fast computation (Parodi and Bortfeld, 2006; Miyatake et al., 2011; Priegnitz et al., 2011; Remmele et al., 2011), the clinical investigations reported so far have relied on full-blown Monte Carlo (MC) simulations of the interaction between the primary beam and the irradiated tissue. In particular, different approaches have been pursued to address the two main steps of modeling the irradiation-induced positron emitter yield and the subsequent imaging process in dependence of the considered ion species and available PET instrumentation, as addressed in the following sections.

17.2 Monte Carlo Modeling of Ion-Induced $\beta^+$-Emitters

Reliable reproduction of the spatial distribution and quantitative yield of irradiation-induced positron emitters requires correct handling of several physical quantities and processes. These comprise the initial characterization of the therapeutic beam and its slowing down in the traversed medium including not only the dominant electromagnetic interactions but also the nuclear processes that may result in positron emitter production, besides causing attenuation of the primary beam and build-up of secondary radiation in depth. While some of these aspects are common to the problem of MC dosimetric calculations in ion beam therapy (cf. Chapter 14), it should be noticed that the partial reaction cross sections yielding positron emitters seldom exceed values of 200 mb (Figure 17.1). Therefore, whereas in terms of nuclear processes the dosimetric calculations are mostly sensitive to a realistic reproduction of the total reaction cross sections and of the main yield of secondary charged hadrons, the PET calculations are extremely sensitive to the specific $\beta^+$-activation channels of typically low probability. Moreover, depending on the primary ion species, the pattern of irradiation-induced activity either is entirely ascribed to positron-emitting target fragments or exhibits a peaked contribution from $\beta^+$-emitting projectile fragments, superimposed onto a pedestal of the activated target fragments (Figure 17.2). Especially the latter remarkable difference in the origin of the activity signal historically resulted in a quite different approach for MC calculations of ion-induced $\beta^+$-emitters, as separately reviewed in the following sections for proton and carbon ion therapy.

17.2.1 Proton Therapy

In the case of primary proton beams, the treatment-induced activity is limited to the target fragments of the irradiated tissue. Indeed, only a full MC simulation including generation and transport of all secondary charged and uncharged particles would be capable of accounting for all possible $\beta^+$-activity production channels. However, it has been estimated in Parodi et al. (2007a) that the positron emitter yield from secondary particles except protons is negligible (<1.5%) even when including the additional neutron background from passively shaped treatment fields. Moreover, although all the major general-purpose MC codes (e.g., Geant4 (Agostinelli et al., 2003), FLUKA (Ferrari et al., 2005; Battistoni et al., 2007), MCNPX (Briesmeister, 1997), and PHITS (Niita et al., 2006)) are capable of predicting proton-induced activation using their own internal nuclear models, it has been shown in Seravalli et al. (2012) that the resulting positron emitter production may considerably differ in terms of both quantitative yield and spatial distribution.
A more accurate approach proposed in Parodi et al. (2002, 2007a,b) and already clinically applied in Parodi et al. (2007c) makes use of available experimental or evaluated cross sections $\sigma_{x\rightarrow Y}$ combined during runtime with the proton fluence $\Phi$. According to this formalism, the total amount $N_Y$ of positron emitters $Y$ produced in nuclear interaction with a nucleus $X$ in a detecting volume $\Delta V$ can be obtained from integration over the primary and secondary proton energy spectrum as

$$N_Y = \int \frac{d\Phi(E)}{dE} f_X \rho N_0 A_X \sigma_{X\rightarrow Y}(E) \Delta V \, dE$$  (17.1)

where $\rho$ is the medium density, $f_X$ is the fraction by weight of the considered nucleus $X$ of atomic weight $A_X$, and $N_0$ is the Avogadro number. Depending on the considered cross-section data, correction for the isotopic abundance of the target nucleus $X$ (e.g., 96.94% $^{40}$Ca in natural Ca) may be required (Parodi et al., 2007a). When performing calculations for clinical cases, the less straightforward information on the elemental tissue composition can be deduced from a proper stoichiometric calibration of the CT data, which are typically used for the patient model in the treatment planning system and in the MC input geometry (cf. Chapter 14), for example, relying on the proposal of Schneider et al. (2000).

Examples of external reaction cross sections used within the FLUKA MC code according to Equation 17.1 for application to PET imaging during and after proton irradiation of phantoms (Parodi et al., 2002, 2005, 2007a,b) and patients (Parodi et al., 2007c) at the Helmholtzzentrum für Schwerionenforschung (GSI) in Darmstadt and Massachusetts General Hospital (MGH) in Boston, respectively, are shown in Figure 17.3. In human tissue, according to the main elemental abundance and the magnitude of the reaction cross sections, the most frequently formed isotopes are $^{11}$C (half-life $T_{1/2} = 20.39$ min), $^{15}$O ($T_{1/2} = 2.03$ min), $^{14}$N ($T_{1/2} = 9.97$ min), and $^{38}$K ($T_{1/2} = 7.63$ min). An example of their calculated spatial distribution is shown in Figure 17.4 for a patient suffering from a clivus chordoma tumor, when taking into account only one of the two orthogonal treatment fields delivering a planned dose of about 0.9 Gy each (Parodi et al., 2007a). It is evident that in addition to the cross-section dependence, the pattern of activated target fragments reflects the estimated tissue composition with dominant carbon abundance and thus $^{11}$C production in fatty tissue, while prevalent oxygen content with related high $^{15}$O yield in soft tissue and calcium content with resulting $^{40}$K activation in bone. As statistical fluctuations are reduced in this approach convolving fluence and cross sections at the voxel level with respect to a full MC computation explicitly simulating pointwise positron emitter production (Parodi et al., 2007a), sufficient counting statistics can be achieved within few hours when using several parallel runs on a high-performance computer cluster.

The final contribution of the simulated $\beta^+$-emitter distributions to the measurable activity signal varies in dependence of the course of irradiation and PET imaging, due to differences up to one order of magnitude in the half-life of the produced isotopes as well as in the timing of in-beam, in-room, or offline data acquisition. Therefore, for accurate PET-based treatment verification, it is strongly recommended to validate the MC prediction of positron emitter yields for the reaction channels that are most relevant to the considered imaging application via dedicated measurements in phantom of known composition (Parodi et al., 2002, 2005; Miyatake et al., 2011). If necessary, the used cross-section data could be tuned to better reproduce the measurements acquired at the own facility. For this purpose, the approach based on Equation 17.1 is indeed more straightforward than relying on the often not accessible internal nuclear models of the codes. The remaining steps for simulation of the irradiation-induced $\beta^+$-activity starting from the predicted positron emission map will be addressed in Section 17.3.

### 17.2.2 Carbon Ion Therapy

In the case of primary carbon ion beams, the treatment-induced activity includes contributions from both projectile...
Monte Carlo Techniques in Radiation Therapy

and target fragments. Therefore, the simplified approach proposed in Equation 17.1 for calculation of proton-induced target fragmentation is no longer applicable, and reliable internal nuclear models are required to properly account for the alteration of the primary beam quality and the generation and transport of the secondary charged fragments, which give a nonnegligible contribution to the β+-activity production. Nevertheless, a simplification has been suggested in Pönsich et al. (2004) for clinical routine application to in-beam PET monitoring of scanned carbon ion therapy, to replace the otherwise too time-consuming full MC simulation approach. The basic underlying idea is that the most relevant information on the in vivo beam range mainly arises from the spatial distribution of the positron-emitting projectile fragments (cf. Figure 17.2), which is less sensitive to the specific tissue stoichiometry in comparison to the target fragment production. Therefore, the prediction of the irradiation-induced positron emitter yield can be split into a two-step process, with a one-time patient-independent generation of a positron emitter database in a reference homogeneous medium followed by

FIGURE 17.3 Example of cross-section values (lines) interpolated from experimental or evaluated data (symbols) and used in the general-purpose MC code FLUKA for calculation of positron emitter production according to Equation 17.1. (Adapted from Parodi, K et al. Phys Med Biol 2007a;52:3369–3387; Parodi, K, Enghardt, W, and Haberer, T. Phys Med Biol 2002;47:21–36; Parodi, K. On the feasibility of dose quantification with in-beam PET data in radiotherapy with 12C and proton beams, PhD thesis, Dresden University of Technology, in Forschungszentrum Rossendorf Wiss-Techn-Ber FZR-415, 2004. With permission.)
a patient- and fraction-specific adaptation of the generated database to the daily treatment and imaging situation. In this way, the time-consuming detailed simulation of the beam transport and interaction has been performed only once in a homogeneous target (PMMA, C₅H₈O₂) containing the most relevant elements of human tissue for pencil-like carbon ion beams at each available therapeutic energy. An example is shown in Figure 17.5. The specific implementation has relied on a dedicated condensed history MC code described in detail in Pönisch et al. (2004) and Hasch (1996). In particular, the more complex handling of nuclear processes has been based on the parametrization of total and partial reaction cross sections proposed in Sihver et al. (1993) and Silberberg and Tsao (1973) and validated against experimental measurements of irradiation-induced activation.

The second step relies on a fast ray-tracing pencil-beam propagation through the patient CT with the same stretching in depth as implemented in the treatment planning system (Kramer et al., 2000). Positron emitter production is obtained via a random MC sampling from the generated database in PMMA repeated as many times as in proportion to the number of delivered ions for each considered energy, adjusting the longitudinal position to the estimated penetration depth in the patient and scaling the lateral position to the actual beam dimension. Moreover, an approximate correction is applied to the target fragment yields to better account for the patient tissue stoichiometry in comparison to PMMA. This correction consists in enhancing the ¹⁵O to ¹¹C target fragment ratio of the database to account for the average higher content of oxygen than carbon in real tissue with respect to PMMA, and in

FIGURE 17.4  (See color insert.) FLUKA MC calculated dose deposition (upper left panel, color bar in units of mGy) and corresponding positron emitter yields (¹⁵O in the upper right panel, ¹¹C and ³⁸K in the lower left and right panels, respectively, color bars in units of produced isotopes for the prescribed dose) for a lateral proton field irradiating a clivus chordoma. The black–white color bar represents the CT number map arbitrary rescaled for display purposes. (Adapted from Parodi, K et al. Phys Med Biol 2007a;52:3369–3387. With permission.)
Monte Carlo Techniques in Radiation Therapy

17.3 Monte Carlo Modeling of Imaging Process

Two approaches have been pursued so far to account for the imaging process in clinical applications of PET monitoring, mostly dependent on the used detector instrumentation regardless of the method followed for calculation of positron emitter production.

The dual-head configuration of the dedicated in-beam PET installation at GSI Darmstadt is known to introduce limited-angle reconstruction artifacts, which become particularly pronounced when moving away from the center of the imaging field of view. Therefore, a trustful comparison of measured and simulated activity distributions can only be achieved when processing the simulated data with the same software used for the reconstruction of the measured data. This requires the explicit simulation of all steps of the imaging process, from the radioactive decay of the positron emitters to the moderation of the emitted positron until its annihilation into two opposed gamma rays, which are then propagated through the surrounding medium up to the coincident detection by the PET scanner. All these steps have been implemented in the same dedicated MC code of Pönisch et al. (2004), starting from the positron emitter yield in the patient CT as obtained from the described adaptation of the initial database in PMMA (cf. previous section). The whole process takes into account the recorded fraction-specific time course of irradiation and imaging as follows. The production time of the positron emitters is identified with the time at which the considered pencil-beam is applied in the dose delivery. Starting from this time, random radioactive decays are simulated according to the isotope half-lives, but only those potentially leading to coincident detection during the registered time course of the PET data acquisition are considered for further processing. The emission and moderation of the positron in the medium are performed starting from the random sampling of the positron range in water from an isotope-specific look-up table. A random direction is then selected and the positron is tracked until the stopping point, as determined by properly stretching the range in water to the patient anatomy according to the CT information. Unless the stopping position is in air (within or outside the patient), the emission of the annihilation photons is explicitly simulated taking into account their angular distribution probability. The propagation of the annihilation photons is then performed until they reach the detector or escape the simulation space, accounting for their main physical interactions (Rayleigh and Compton scattering) according to look-up tables and including established variance reduction techniques to increase computational efficiency (Pönisch et al., 2004). This is particularly important taking into account the extremely low geometrical acceptance of the detector, covering only about 9% of the entire solid angle. No explicit interaction with the crystal is simulated, but acceptance of a valid hit on the
detector is decided by considering the energy threshold of the positron camera and the energy resolution of the BGO (bismuth germanate) scintillation blocks. The final output is a list mode of detected events in the same format as the measured data, allowing for the usage of the same reconstruction algorithms. An example of experimental validation is shown in Figure 17.6a. With this approach, consistent datasets have been compared throughout the entire pilot project at GSI, enabling clinically useful information to be inferred from the in-beam PET monitoring (Enghardt et al., 2004a,b). According to Pönisch et al. (2004), a simulated PET list mode dataset could be obtained in less than 5 min on a single computer for a typical patient treatment, allowing the results of the PET treatment monitoring to be available within about 30 min after completing the therapeutic irradiation.

More recently, the latter step of the dedicated code (Pönisch et al., 2004) accounting for the response of the GSI in-beam PET detector (in terms of geometrical and energy acceptance) has been coupled to the output of the general-purpose code FLUKA for simulation of the entire PET imaging process in phantom experiments with carbon and oxygen ion beams. In particular, FLUKA was successfully used for the transport of the ions, the creation of the β+-active nuclei, the decay, the electron–positron annihilation, and the transport of the annihilation photons until escape from the irradiated target (Sommerer et al., 2009). For this study, a new variance reduction technique was introduced in the code for directional biasing of the annihilation photons toward the detector of very poor geometrical coverage, complementing useful available tools such as replication of the β+-decay from the produced isotopes to enhance detection in the considered measuring time window. Despite the reported considerable gain in computational efficiency up to a factor of 10 for the direction biasing and 25 for the decay replication, the execution times were deemed to be still too long for clinical routine application. Nevertheless, this work represented an important first step toward the usage of a general-purpose code for full MC simulation of both the irradiation and imaging processes in PET-based verification of heavy ion therapy.

In the case of full-ring PET detector geometries as used at MGH Boston and HIT Heidelberg, the considerably improved coverage of the solid angle significantly reduces imaging artifacts in comparison to limited-angle dual-head configurations. Therefore, an easier approximation of the combined effects of the image formation and reconstruction process consists in a Gaussian smoothing of the MC simulated positron emitter distributions, properly weighted by isotope-dependent factors that account for the number of decays contributing to the activity signal in the considered acquisition time window (Parodi et al., 2007b,c). The parameters of the Gaussian point spread function can be determined from the specifications of the scanner or be empirically obtained from point source measurements. An example of experimental validation is shown in Figure 17.6b. This simple and very fast approach has been already successfully used for several clinical investigations of PET imaging in proton therapy, where it has been applied to CT-based positron emitter

![FIGURE 17.6](image-url)

(a) Depth distributions of β+-activity induced by a 212 AMeV carbon ion beam in a thick PMMA target reconstructed by means of backprojection from measured (thick solid line) and simulated (Pönisch et al., 2004) (dotted line) list mode data acquired in about 5 min long irradiation. The calculated contribution of the most abundant projectile and target fragments is additionally shown by the thin lines of different gray scale. (b) Depth profiles of in-beam PET activity induced by a 10 min irradiation of a thick PMMA target at 140 MeV proton beam energy. The simulation (dotted line) obtained from the Gaussian smoothing of the positron emitter yield calculated with FLUKA on the basis of Equation 17.1 and the cross-section data of Figure 17.3 is compared to the measurement (solid line). The dashed line indicates the expected range. Only relative activity distributions are shown as the in-beam PET camera at GSI could not provide absolute reconstruction. In this example, limited-angle artifacts are not an issue due to the imaging of a compact activity distribution in the central plane of the scanner.

yields generated by full MC codes such as FLUKA (Parodi et al., 2007c) and Geant4 (Zhu et al., 2011).

Indeed, dedicated MC-based tools that were originally developed for characterization of nuclear medicine instrumentation could be coupled to the output of the general-purpose codes for full MC simulation of the imaging process, including the detailed handling of photon interactions in the detector crystals for a complete modeling of the detector response. In particular, the Geant4-based simulation platform GATE (Geant4 Application for Emission Tomography), which was initially developed for research on PET as well as single-photon emission
tomography (SPECT), has been recently extended to radiation therapy applications, including in-beam PET modeling for hadron therapy (Jan et al., 2011). However, besides the unavoidable increase of computational time, the feasibility of such a full MC handling requires detailed knowledge of the PET scanner and of its proprietary data formats as well as applicability of the reconstruction software to the simulated data, which is often not available/possible for commercial PET instrumentation. Moreover, it is still controversial whether the realistic addition of statistical noise in the full MC simulation, which is not accounted for by the smoothing approach, is a real benefit for the purpose of PET-based treatment verification due to the resulting degradation of image quality in the simulated image used as reference.

17.4 Remaining Challenges for Clinical Implementation

As reported in the first clinical investigations of in-beam, in-room, and offline PET monitoring of carbon ion and proton therapy (Enghardt et al., 2004a,b; Parodi et al., 2007c; Knopf et al., 2009; Zhu et al., 2011), the main factors for a correct prediction of the irradiation-induced activity are the knowledge of the reaction cross sections and of the tissue composition in the modeling of ion-induced β+-emitters (cf. Section 17.2), as well as the determination of the contribution to the measurable activity in the imaging process (cf. Section 17.3). Indeed, the modeling of ion transport and nuclear processes resulting in β+-activation together with the explicit or approximate description of the image formation process can be extensively validated and possibly fine tuned via irradiation and PET imaging of phantoms of known composition. However, challenges remain in clinical application for the determination of the elemental tissue composition and the modeling of the so-called biological washout, which is responsible for the fact that part of the activity produced in living tissue is washed away by physiological processes such as perfusion.

While the CT-based stoichiometric calibration proposed in Schneider et al. (2000) is typically sufficient for reliable dosimetric calculations (Espana and Paganetti, 2010) (cf. Chapter 14), it may miss details of the relative elemental composition for organs sharing the same radiological properties but exhibiting significant differences in the oxygen and carbon abundance. Especially in the case of proton therapy where the irradiation-induced activity is entirely ascribed to target fragments, a failure in the identification of the elemental composition results in an erroneous modeling of the activity not only in terms of signal strength but also beam range (as deduced from the distal activity fall-off) due to the different energy thresholds of the different activation channels (cf. Figure 17.3). In fact, changes in the carbon-to-oxygen ratio have been shown to result in almost 2 mm differences in the estimated proton range (Parodi et al., 2005), while range variations from 1 to 5 mm were reported in the MC studies of Espana and Paganetti (2010) and Espana et al. (2011) addressing the influence of different CT stoichiometric calibration in combination with different reaction cross sections for different implementations of PET-based monitoring of proton therapy. In clinical application of offline PET imaging of proton therapy, observable discrepancies between the calculated and the measured activity (predominantly from 11C) were reported and ascribed to an erroneous estimation of carbon abundance in tissue such as bone marrow (Knopf et al., 2009) and white brain matter (Parodi et al., 2007), which show remarkable deviation from the general CT-based calibration of Schneider et al. (2000).

A possibility to improve tissue classification using complementary patient-specific information such as magnetic resonance imaging (MRI) has been recently reported (Parodi et al., 2011). For carbon ion therapy, the influence of tissue stoichiometry is less pronounced due to the major contribution from positron-emitting projectile fragments, especially in terms of range verification (Pönisch et al., 2004).

Regardless of the primary ion species, considerable differences between measured and calculated activity may occur due to the physiological processes carrying part of the formed isotopes even far away from the place of production. The relevance of these dynamic processes increases with the time distance between irradiation and imaging, introducing distortions of the physically formed activity that are especially pronounced in offline PET imaging (Figure 17.7). This justifies why wash-out processes could be in first approximation neglected in the clinical application of in-beam PET imaging of carbon ion treatments at GSI (Enghardt et al., 2004a,b), but had to be considered in the clinical investigations of in-room and offline PET-based verification of proton and carbon ion therapy at MGH (Parodi et al., 2007; Knopf et al., 2009; Zhu et al., 2011) and HIT (Parodi et al., 2011), respectively.

The effects of the biological washout can be included in the modeling of the image formation process, converting the simulated positron emitter yield into the expected β+-activity distribution. As the hot chemistry taking place immediately after the formation of β+-active isotopes in living tissue is very complex and not yet completely understood, compartment models well established in nuclear medicine are not applicable to low-statistics PET imaging of ion therapy. Nevertheless, the washout effects can be accounted for via the introduction of a biological decay in addition to the physical one, as initially suggested in Mizuno et al. (2003) on the basis of animal studies with implanted radioactive ions. For application to patient cases, two different approaches have been proposed in the literature for offline PET verification of proton therapy and in-beam PET monitoring of carbon ion therapy, respectively, both neglecting differences in the biological pathways of the different formed isotopes. The first approach (Parodi et al., 2007c) accounted for tissue-specific biological parameters in the postsimulation weighting of the MC-produced positron emitter yields contributing to the calculated activity in the considered time window (Figure 17.7). The second one (Fiedler et al., 2008) introduced an empirically determined dose-dependent effective half-life to correct the number of decays to be explicitly simulated as potentially giving rise to a detectable signal in the in-beam PET acquisition (Figure 17.8),
Despite the encouraging qualitative results of both approaches illustrated by Figures 17.7 and 17.8, quantitative differences in the simulated activity distributions up to 30% in intensity and 4 mm in range were reported for offline PET imaging of proton therapy in anatomical regions especially subject to perfusion (Parodi et al., 2007; Knopf et al., 2009). Therefore, work is still ongoing at several institutions involved in clinical investigations of PET monitoring to improve the washout modeling, trying to gain patient-specific information from additional imaging modalities. However, these activities go beyond the topic of MC modeling of radiation transport and interaction addressed in this book.

17.5 Conclusion and Outlook

This chapter has reviewed the main implementations and issues of MC calculations for PET-based treatment verification of proton and carbon ion therapy. In particular, it has shown that different computational approaches have been historically pursued, with different approximations introduced in dependence of the considered ion type and PET instrumentation for the sake of fast execution time. Encouraging results could be so far demonstrated in several pilot investigations of in-beam, in-room, and offline PET monitoring of carbon ion and proton
therapy at GSI, MGH, and HIT (Enghardt et al., 2004a,b; Parodi et al., 2007c, 2011; Zhu et al., 2011). Moreover, several ongoing research activities are focused toward the establishment of computationally efficient full MC simulation approaches mostly relying on general-purpose codes being thoroughly validated against available experimental measurements (Seravalli et al., 2012; Sommerer et al., 2009; Jan et al., 2011). In this respect, new experimental campaigns of $\beta^+$-activation in phantoms of known composition would be highly beneficial (Seravalli et al., 2012; España et al., 2011). However, despite the promising results and ongoing developments, major challenges remain for clinical application that are mostly related to the knowledge of the tissue elemental composition and physiological properties. The former aspect is especially important for the modeling of $\beta^+$-activity distributions limited to target fragments as in proton therapy. The latter issue of biological washout currently poses the most serious limitation to the indications which may benefit from PET monitoring especially in data acquisition taken long after irradiation, thus motivating new efforts for dedicated in-beam or in-room instrumentation (Shakirin et al., 2011) as well as the search for alternative range monitoring techniques such as real-time prompt gamma imaging or ion radiography before or in between treatment (cf. Chapter 18).

References


Nuclear Reaction Data Center network which is coordinated by the IAEA Nuclear Data Section, 2000. Available at http://www.nndc.bnl.gov/exfor/exfor00.htm.


18.1 State of the Art in Prompt Gamma Research

In addition to the previously discussed methods for using positron emission tomography (PET) imaging techniques to measure induced positron annihilation gamma emission in the patient, researchers have also studied imaging prompt gamma (PG) emission from the patient, as a means of verifying proton and ion beam treatment delivery. During proton or heavy ion therapy, primary particles from the treatment beam interact with elemental nuclei in the tissue, which can leave behind an intact nucleus in an excited state. This excited state quickly (<10^{-9} s) decays by emitting a characteristic PG. Since the PG emission only occurs in regions in which the treatment beam interacts in the tissue, the distribution of PG creation in the patient is correlated to the distribution of dose delivery to the patient. Also, since the excited nuclear states are quantized, each element emits a unique and characteristic PG spectrum. Thus, spectral analysis of PG emitted from tissue may provide useful information about the elemental concentration, composition, and type of the tissues irradiated during treatment.

The possible applications of treatment delivery verification and tissue composition analysis have led to an increasing number of studies, particularly Monte Carlo (MC) studies, of PG emission. In fact, many researchers have made use of the specific capabilities of many MC codes such as MCNPX (Waters et al., 2005), Geant4 (Agostinelli et al., 2003), and FLUKA (Fasso et al., 2003) to study PG emission during proton beam and heavy ion beam therapy. MC has proven to be a very powerful tool for these studies, as it has allowed the users to study a wide range of topics, including characterization of PG emission from proton-irradiated phantoms and tissues, detector design, image reconstruction, and clinical applications such as proton beam range verification and tissue identification.

18.2 Basic Properties of PG Emission

An important first step in the development of PG imaging techniques is the study and characterization of the basic properties of PG emission. MC is particularly well suited for these studies due to the availability of a wide range of physics models and data cross-section tables. MC also offers a large degree of flexibility for tracking and recording the characteristics of emitted PG gammas.

For this purpose, Polf et al. (2009a) used the MCNPX MC code to perform an initial investigation of PG emission from biological tissues undergoing proton irradiation to investigate its potential use for in vivo analysis of irradiated tissues. For this study, small tissue phantoms were defined within the model and irradiated with therapeutic proton beams. The spectra of
emitted PG gammas from several different tissue types and several of the major elemental constituents were recorded as a function of position around the phantom. The physics defined in these simulations included energy straggling, multiple Coulomb scattering, and elastic and nonelastic nuclear interactions (including nuclear deexcitation processes). This study showed significant differences in the PG spectra emitted from muscle, lung, adipose, and bone tissues due to the differences in their composition and density. Also, the dependence of the PG emission intensity (production cross section) as a function of proton beam energy was calculated. These MC calculations indicated that a maximum PG emission cross section for all tissue types studied occurred for proton beam energies ranging from 20 to 30 MeV, with the PG production steadily decreasing as the beam energy approached the highest energies (~250 MeV) currently used in clinical proton therapy.

18.3 Range Verification and Tissue Composition Analysis

The main clinical application of PG emission studied to date has been its use in verifying the depth of the proton Bragg peak within the patient during treatment. Many initial studies focused on the correlation between the range of PG production in water and tissue phantoms and the proton beam Bragg peak using MC calculations (Min et al., 2006; Feng et al., 2007, 2010; Kim et al., 2009; Polf et al., 2009a; Le Foulher et al., 2010). In these studies, the researchers have used many flavors of MC, including MCNPX, FLUKA, and Geant4 to track and tally the location within the phantom at which the PG gamma was created and record these data in 1D, 2D, or 3D histogram files. The MC calculated PG emission distributions were compared to the MC calculated distributions of delivered dose. Results of these studies in tissue phantoms have all shown that the range of PG production and the proton beam dose distribution are strongly correlated, with the peak PG emission occurring ~2–4 mm proximal to the Bragg peak.

In an MC study using several clinical cases, Moteabbed et al. (2011) used Geant4 to study the correlations between PG production, annihilation gamma production (used in posttreatment PET imaging), and the dose delivered within the patient during treatment delivery. For these studies, the patient CT data as well as all pertinent information for treatment delivery were imported from a clinical treatment planning system into an MC model of a proton therapy delivery nozzle. MC calculations were performed to simulate the patient treatment delivery and both the prompt and annihilation gamma yields within the patient were tallied. Results from these studies show a strong correlation between PG production and the in vivo dose distribution for a wide range of clinical proton therapy cases. As an example, the calculated data for a prostate cancer case is shown in Figure 18.1. For this case, the range and lateral extent of the PG production in the patient correlates well with the distribution of dose delivered by the treatment beam. Along with the PG–dose correlation, the study by Moteabbed et al. showed that the intensity of the PG production was strongly dependent on the density of the tissue (i.e., bone, muscle, and lung) and the presence of large heterogeneities such as bone–air interfaces. Results of this study helped to show the feasibility of using PG emission to verify in vivo proton beam range in a clinical setting.

In addition to range verification, Polf et al. (2011) used the MCNPX MC code to study the potential for using PG emission to determine the concentration of the major elemental constituents of irradiated tissues. In this study, an MC model of a clinical proton therapy beam irradiating a large tumor within a patient was used. The CT dataset of a patient with a large lung tumor was loaded into the MC model with the material definition and density of the patient’s normal tissue and tumor being defined according to their ICRU definitions (ICRU, 1993) and the tumor defined as ICRU average soft tissue. To simulate compositional changes due to tumor hypoxia, the percent concentration of oxygen within the tumor was then systematically reduced, by adjusting the material definition card in the MCNPX input file accordingly, over a series of MC calculations. For each MC calculation, at a given percent oxygen concentration in the tumor, the characteristic PG emission from oxygen produced in the tumor was tallied. From these calculations, it was shown that the production of PG gammas from oxygen decreased as a logarithmic function of oxygen concentration within the tumor.

FIGURE 18.1  (a) Proton treatment plan dose on a CT image for a patient treated for prostate cancer. (b) PG emission overlaid on the same CT image. Comparison of the dose and PG distribution shows a strong correlation between dose and PG emission. (Adapted from Moteabbed, M., S. Espana, and H. Paganetti, Phys. Med. Biol., 2011. 59: 1063–1082.)
18.4 Detector Design Studies

Favorable results from preliminary MC studies of PG production in irradiated tissues have led many researchers to study the design of efficient PG detectors for use in clinical proton and heavy ion therapy applications. Since characteristic PGs emitted from tissue range in energy from 2 to 15 MeV, current gamma ray detectors (mostly designed for lower-energy applications) suffer from low detection efficiency. Since the total number of PGs emitted during radiotherapy is limited due to short patient irradiation times, these low efficiency detectors have not shown the ability to measure enough PG emission data to produce statistically relevant and useful images. In addition, the higher energy of the PG emission makes standard physical collimators used for low-energy medical and nuclear imaging applications ineffective for encoding spatial information about the PG emission within the measured data necessary for reconstructing images. Therefore, a large effort has recently been put forth to design efficient and effective detectors capable of measuring spatial and initial energy information of PG emission from tissue under clinical conditions.

Many recent MC studies have focused on the design and development of detection systems specifically for measuring and imaging PG emission from tissue during proton and ion beam therapy. This includes studies simulating a number of different designs of two-stage (Feng et al., 2008; Kang and Kim, 2009; Frandes et al., 2010; Roellinghoff et al., 2011) and three-stage Compton cameras (Richard et al., 2009; Peterson et al., 2010; Robertson et al., 2011), as well as pinhole and slit cameras (Min et al., 2006; Kim et al., 2009; Bom et al., 2012), and multislit collimated array detectors (Min et al., 2012) using a wide range of materials and detector configurations. The main purposes of these studies were to optimize the PG detection efficiency of the modeled detectors.

For example, Figure 18.2 shows the designs for two different proposed PG detection systems, a three-stage Compton camera and a slit camera, studied by Peterson et al. (2010) and Bom et al. (2012), respectively, using the Geant4 MC toolkit. For these studies, MC calculations were used to determine the feasibility of the designed PG detection systems.
of detecting an adequate PG signal under clinical beam delivery conditions to allow for determination of the proton beam range. This was accomplished by using MC simulations to track PGs emitted from the irradiated tissue phantom as they propagate through the modeled detectors and count the number of individual interactions for each PG, as well as record the position and energy deposition of each interaction. For the Compton camera (Figure 18.2b), Peterson et al. performed a series of MC calculations in which they varied the width and thickness of each stage, as well as the spacing between each stage of the camera to determine the specific design that optimized the detected signal from characteristic oxygen (6.13 MeV), carbon (4.44 MeV), and nitrogen (2.31 MeV) PGs. In contrast, Bom et al. used MC to study and optimize the energy deposition thresholds used to trigger a measurement in their detector, which would allow them to discriminate between incident PG emission and secondary neutrons and scattered protons and gammas that interact with the detector during irradiation. From the results of these MC studies, both Peterson et al. and Bom et al. were able to determine optimal design characteristics of their detectors. Using these optimal detector designs, they were then able to calculate the overall detection efficiency of the modeled detectors under clinical proton beam delivery conditions, and thus make initial conclusions as to their feasibility for use in proton beam range verification.

18.5 Image Reconstruction Studies

Additionally, the PG interaction data from the MC calculations have been used by several researchers to study methods of reconstructing images of the PG emission from irradiated phantoms. This includes study of the gamma ray tracking (GRT) technique (Deleplanque et al., 1999; Schmind et al., 1999) and stochastic origin ensemble (SOE) technique (Sitek et al., 2008; Andreyev et al., 2011) used in conjunction with MC calculated data from a Compton camera, and with filtered back projection (FBP) used with MC calculated data from a collimated gamma camera (similar to clinical single-photon emission computed tomography systems). With the MC modeling, the physical characteristics of the detectors can be simulated, including pixel size, Doppler broadening, energy resolution, and polarization effects. The effects these characteristics have on the achievable spatial resolution and noise levels of PG measurements and the resulting effects on images reconstructed with GRT (Feng et al., 2008; Richard et al., 2009; Frandes et al., 2010) and FBP (Park et al., 2010) have been studied, by varying the their magnitude over a series of MC calculations.

Additionally, simulated PG data from a modeled three-stage Compton camera were used by Mackin et al. (2012) as input data for design/development studies of a new reconstruction method tailored specifically for PG image reconstruction. In this study, the SOE algorithm originally defined by Sitek (2008) for detection of lower-energy (<1 MeV) gammas was adapted to work with the specific characteristics of PG emission and detection (with a Compton camera) during proton radiotherapy. The data files output by the MC model were formatted to mimic the “list-mode” data file output by several experimental Compton cameras. The MC data files contained a list of the position and energy deposition of each PG interaction in each stage of the camera. The MC data was then used in lieu of actual measured data as input data for the SOE reconstruction code. Images of the PG emission reconstructed using the SOE code with input data calculated by the MC model are shown in Figure 18.3. In addition to recording the PG interaction positions in the camera, the MC model also recorded the positions where each PG was emitted (Figure 18.3d). The quality of the images produced by the SOE algorithm from the CC data was quantified by comparing the images to the distribution of the recorded emission positions. In this manner, the effect of each SOE design parameter on the quality of the reconstructed images could be directly tested during each phase of the algorithm development.

18.6 Validation of Monte Carlo-Based PG Calculations

An important consideration for the use of MC in the study of PG emission applications in radiotherapy is the proper tuning of all parameters of the MC model (including the model geometry, material definitions, particle tracking, and physics models) to ensure that the calculated data can accurately reproduce the physical processes occurring during treatment delivery. This validation process is usually performed through the direct comparison of calculated results from the MC model and
measured data. Unfortunately, validation studies of MC models used for PG emission studies have been slow to be published. This is most likely due to the scarcity of beamtime available for experimental measurements at proton and carbon ion therapy facilities, leading to a lack of direct measured data of total PG emission rates and PG spectra for use in MC validation studies.

In one such study, Polf et al. (2009b) performed validation studies of an MC model used for calculating the spectrum of PGs emitted from tissue phantoms during proton irradiation. For this study, an MC model of an experimental setup consisting of a shielded high-purity germanium detector setup to measure PG emission from small plastic phantoms irradiated with a proton pencil beam was developed using Geant4. The MC calculations included physics models for gamma interactions ranging from 250 eV to 10 TeV, and both elastic and nonelastic proton–nuclear interactions (including an intranuclear cascade deexcitation model). With this MC model, Polf et al. (2009b) were able to reproduce the measured PG spectra from irradiated soft tissue and bone equivalent plastic phantoms, as shown in Figure 18.4. This included the presence of the measured PG emission lines from the main elemental constituents of the phantoms, namely, oxygen, carbon, and calcium. However, the nuclear scattering models (at the time of this study) were not able to correctly calculate the nuclear Doppler broadening of the PG emission lines present in the measured spectra as seen in the 4.44 MeV emission line from carbon in Figure 18.4.

In another validation study, Le Foulher et al. (2010) compared measured PG count rates and spectral emission from plastic and water phantoms during carbon ion irradiation to calculations from an MC model. In this study, the researchers used Geant4-based MC model that recreated the geometry of the experimental setup (including detectors, shielding, and phantoms) used for the measurements, as well as the parameters of the carbon ion beam used for the irradiations. The MC model included all standard electromagnetic processes that dictated the PG propagation and interactions, as well as nuclear physics packages to model interaction processes, including nuclear scattering, fragmentation, and deexcitation of the excited nucleus.

The results of these studies showed that the nuclear physics packages used in their model greatly overestimated the PG production in the phantoms during carbon ion irradiation. However, their investigation and comparison of the fragmentation cross sections within the nuclear physics models showed very good agreement with experimental measurements. Therefore, the discrepancies in the measured and MC calculated PG count rates were attributed to uncertainties in the gamma-ray multiplicity values contained within the nuclear models. These results highlight the importance of continued testing and refinement of the MC physics models to ensure accurate and usable results for MC studies of PG emission during proton and ion beam therapy.

18.7 State of the Art of Proton Radiography

In 1946, Robert Wilson (Wilson 1946) recognized the potential for proton beam therapy due the inverse relationship between the ionization energy loss and the energy of a proton beam passing through tissue. Subsequently, proton radiography was first investigated as a radiologic tool in the early 1970s and 1980s (Steward and Koehler, 1973; Hanson et al., 1982). However, the success of x-ray radiography, which quickly became the most common cancer diagnostic tool in hospitals, rapidly overshadowed proton radiography. Nevertheless, despite a difficult start, proton radiotherapy, as well as heavy ion therapy, is becoming more and more attractive to the radiation oncology community as technology improves. Indeed, protons present a finite range in tissue that can allow for better tumor conformality and therefore better sparing of the surrounding healthy tissue. Although protons’ intrinsic imaging capabilities are limited (Schneider and Pedroni, 1994), by using higher energies than the ones for treatment, ion radiography further takes advantage of the aforementioned plateau region and is able to produce patient setup images while delivering significantly less imaging dose than conventional x-rays (Depauw and Seco, 2011).

18.8 Protons versus X-Ray

The subject of proton versus conventional x-rays has been the object of a multitude of studies (Agostinelli et al., 2003; Ryu et al., 2008; Depauw and Seco, 2011; Depauw et al., 2011), especially with MC as it allows a fair comparison between the two modalities without taking into account the current state of the art in detector technology. Current technology is obviously much more advance on the x-ray side as it benefits from more than a hundred years of study, as well as the broad development of photon radiotherapy centers all over the world. When comparing protons with photons for imaging purposes, it is customary to define two main metrics: spatial resolution and density resolution. In

![FIGURE 18.4 Comparison of measurement (black symbols) to MC calculations (gray lines) of secondary gamma emission spectra from Lucite (upper curves (×10 for display purposes)) and bone equivalent plastic (lower curves) phantoms irradiated with clinical proton beams. (Adapted from Polf, J.C. et al., Phys. Med. Biol., 2009. 54: N519–N527. With permission.)](image-url)
a recent study (Depauw and Seco, 2011) using the MC toolkit GEANT4 (Schneider and Pedroni, 2012), an imaging phantom was specifically designed to properly compare these aforementioned resolutions. Figure 18.5 presents the $30 \times 30 \times 8$ or $15 \text{ cm}^3$ water phantom with 21 rows of cylindrical inserts made of different materials and densities: air ($0.001 \text{ g/cm}^3$), lung type tissue ($0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, \text{ and } 0.8 \text{ g/cm}^3$), adipose tissue (1 g/cm$^3$), soft tissue (1 g/cm$^3$), and bony type tissues inserts (1.1, 1.2, 1.29, 1.39, 1.49, 1.58, 1.67, 1.77, 1.86, and 1.92 g/cm$^3$). For each material, the phantom contains a total of 10 inserts with diameters ranging from 20 to 0.88 mm, decreasing in size. Different simulations considered different phantom thickness (8 or 15 cm), different insert thickness (2, 4, or 8 cm), as well as different incident proton or photon energy (200, 300, 400, and 490 MeV, and 50 keV, 100 keV, 1 MeV, and 2 MeV, respectively). Similar imaging doses were considered across modalities and energies, thus giving an adequate point of reference for an imaging capabilities comparison study. Figure 18.6 presents three radiographs with three modalities for similar delivered doses.

18.9 Proton Radiography Image Reconstruction Differences

The processes for generating radiographs from protons and from photon are very different. Proton radiographs are “energy maps.” In each pixel, the imaged value is the range of the absorbed energy $E_{abs}$ corresponding to the range in water of the mean energy loss of all the protons going through this particular pixel. This mean absorbed energy $E_{abs}$ has to be computed on a proton-by-proton basis, and different techniques are considered to evaluate the path of each particle such as “the most likely path” (Schneider and Pedroni, 1994). When considering a proton-by-proton system, one would be confronted with the problem of secondary protons generated through nuclear interaction in the patient (~1%/cm). These secondary protons present large scattering angles as well as lower energies and therefore constitute an important source of noise, as shown in Figure 18.7. By evaluating radiographs with energy and angular cuts, it is shown that the effect of secondary protons can be reduced. Another recent study suggests that the air gap between the subject and the radiography system would also play an important role in the final image quality that can be achieved (Schulte, 2005). On the other hand, x-ray radiographs are simple “fluence maps”: in each pixel, the imaged value is the number of particles going through this

![Figure 18.5](image1.png)

**FIGURE 18.5** MC-based water phantom filled with rows of 21 different materials with densities ranging from 0.001 g/cm$^3$ (air—#1) to 1.92 g/cm$^3$ (bony material—#21). This phantom was designed to properly compare protons with photons for imaging purposes.

![Figure 18.6](image2.png)

**FIGURE 18.6** Reconstructed phantom radiographs using (a) a 200-MeV proton beam, (b) a 50-keV x-ray beam, and (c) a 1-MeV x-ray beam for similar absorbed dose.

![Figure 18.7](image3.png)

**FIGURE 18.7** Reconstructed proton radiographs with, (a) primary protons only, (b) primary and secondary protons, (c) primary protons with angular and energy cuts, and (d) primary and secondary protons with angular and energetic cuts. This shows the importance of correcting for secondary protons to significantly reduce the noise and enhance the output of a proton radiography system, in a way such as the antisscatter grid in an actual x-ray.
269

Monte Carlo Studies of Prompt Gamma Emission and of Proton Radiography/Proton-CT

pixel. This study also incorporated a 2° angular cut in order to ameliorate the image quality by appropriately simulating the antiscatter grid of an actual x-ray system, and make the comparison more meaningful.

18.10 Spatial Resolution and Density Resolution

The spatial resolution is defined “broadly” as the minimum size in millimeters needed for a specific region of interest to be visually resolvable. It was therefore retrieved as the smallest visible insert for each substudy.

The authors proceeded by computing the contrast-to-noise ratio (CNR) for each insert in each substudy as

$$\text{CNR} = 20 \text{ log} \left( \frac{|f - b|}{\sqrt{\sigma_f^2 + \sigma_b^2}} \right). \quad (18.1)$$

where $f$ is the mean signal value in the region of interest, $b$ is the average background noise, and $\sigma_f$ and $\sigma_b$ are their respective standard deviations. The CNR is commonly used as a “figure of merit” in the characterization of imaging systems. The CNR corresponds to the difference in signal-to-noise ratios (SNR) between two adjacent regions of interest and describes the ability to properly distinguish an area of interest relative to background. In the case of radiotherapy, this would define the ability to distinguish cancerous tissue relative to its surrounding normal tissue.

Figure 18.8 depicts the difference in spatial resolution between protons and photons. It includes results from one substudy with a defined phantom thickness and inserts thickness, with the entire set of both proton and x-ray energies.

18.11 Proton Radiography, kV and MV Image Quality for Identical Integral Dose

The image quality of kV and MV x-ray radiographs and proton radiographs was compared for approximately identical integral absorbed doses in a water phantom, using GEANT4 MC (Agostinelli et al., 2003). Five million protons and 50 millions x-rays were simulated through a similar-sized water tank ($30 \times 30 \times 8$ or $15 \text{ cm}^3$), and the total energy deposited to the volume was recorded. This simulation permitted the assessment of the number of incident particles per modality required to obtain approximately equal absorbed dose. The authors concluded that 50 millions photons at 100 keV deliver approximately the same absorbed dose to the 15-cm-thick phantom as approximately 30,000 200 MeV protons or 55,000 490 MeV protons.

In Figure 18.9, the contrast-to-noise ratio, CNR, is presented for the proton radiography, kV and MV x-ray images obtained with identical integral dose. Protons provide better CNR results than kV and MV for significantly lower integral doses. For low- and high-density regions ($\rho < 0.6 \text{ g/cm}^3$ or $\rho > 1.5 \text{ g/cm}^3$), CNR values are excellent for protons, and acceptable for x-rays. In the medium density region ($0.6 \text{ g/cm}^3 < \rho < 1.5 \text{ g/cm}^3$), MV x-rays yield very poor CNR levels.
18.12 Proton Radiography: A Step toward Proton-CT

By imaging from different angles, proton radiography opens the door to proton computed tomography (pCT). One difficulty then resides in defining a proper reconstruction algorithm, which was investigated through MC simulations in various studies (Schulte et al., 2005; Li et al., 2006). Various reconstruction techniques were investigated: (i) straight line path (SLP), (ii) most likely path (MLP), and (iii) cubic spline path (CSP) using an algebraic reconstruction technique (ART). The straight line path is the simplest (though most naïve) solution of the path the proton travels through any media. It assumes a straight line connecting the entrance and exit points of the proton. The MLP approach identifies the likeliest path based on the proton energy, direction of motion, and the entrance and exit positions. The CSP approach used a cubic spline

![Figure 18.9](image1) **Contrast-to-noise ratio comparison between radiographs with a scanned pencil 200-MeV proton beam, a pure 100-keV x-ray beam, and pure 2-MeV x-ray beams. Each line corresponds to the same ROI area for each material density (legend is cylinder diameter); the hashed area reflects the density range for which no inserts were visible.**

![Figure 18.10](image2) **(a) Phantom used for pCT reconstruction algorithm. It consists of an outer shell with bone density (1), with a water inner filling (2), and strips of either air (black) or bone (white) at different densities: 2 lp/cm (3, 11), 2.5 lp/cm (4, 12), 3 lp/cm (5, 13), 3.5 lp/cm (6, 14), 4 lp/cm (7, 15), 4.5 lp/cm (8, 16), 5 lp/cm (9, 17), and 6 lp/cm (10, 18). (b) Reconstruction of the phantom with the ART algorithm using different proton path estimation approaches.**
Monte Carlo Studies of Prompt Gamma Emission and of Proton Radiography/Proton-CT


fit that is applied to the entrance and exit positions to estimate the proton’s path length. The MLP and CSP approaches utilize ART reconstruction technique, an iterative solution of the projected relative electron density (or proton stopping power) along a single path. The N projections of relative electron density (or proton stopping) are used to generate \( N \times N \)-dimensional reconstructed vector representing an object slice or slices.

Figure 18.10a shows the elliptical phantom used to test the various reconstruction methods, where the line pairs are composed of either air (black) or bone (white). In Figure 18.10b, the reconstructed elliptical phantom is provided using GEANT4 MC calculation algorithm. The SLP provides the worst estimate of the pCT prediction, where a resolution of 2.5 lp/cm was observed. The spatial resolution of the air density and bone density patterns for the MLP and CPS approaches was 5 and 4.5 lp/cm, respectively. The ART reconstruction with either MLP or CSP generated comparable quality pCT images.

References


Fasso, A. et al., *The FLUKA Code: Present Applications and Future Developments*, in CHEP032003, La Jolla, CA, USA.


19.1 Introduction

Monte Carlo simulation has broad application in the field of radiotherapy, beyond the well-known applications in dosimetry and treatment planning. The method allows development and testing of virtual devices using virtual experimentation. This chapter provides a survey of applications of Monte Carlo simulation in the design of treatment devices for conventional linacs used for x-ray and electron therapy. Examples selected from the research, clinical, and consulting experience of the author show the power of the method, and elucidate general principles of treatment device design. The examples are loosely ordered from the exit window of the waveguide, progressing through the different components of the treatment head, through the patient and into the flat panel detector (Figure 19.1).

The EGS4 Monte Carlo system (Nelson et al., 1985) was used for the earlier work, EGSnrc (Kawrakow, 2000), once it was available. The EGS4/EGSnrc user code BEAM (Rogers et al., 1995) was used for treatment head simulation, sometimes with modification, the MCRTP user code (Faddegon et al., 1998) for dose calculation. The results include experimental validation, showing the accuracy of simulation for practical applications. The devices designed have for the most part been fabricated and tested (with the exception of the secondary scattering foils). Many are commercially available.

19.2 Target for Therapy

The accelerated electron beam incident on x-ray targets generates considerable heat. The targets are generally cooled with circulating, chilled water. One design used a gold target cooled by water channeled beneath the target, with water in the beam path. When some of these targets developed hairline cracks, pressurized water would leak out from the cooling channel and drip on components such as the monitor chamber, potentially shorting them out.

Monte Carlo simulation aided in the design of a new target (Figure 19.2). The primary concern was to eliminate the water from the beam path. In addition, the x-ray beam generated in the new target needed to conform closely to the beam from the original target, not just for one beam energy, but for the full range of x-ray beam energies available for treatment on the different accelerator models that used this target. To maintain a comparable dose rate for the same beam current, tungsten was chosen for the target material, being close in atomic number to gold. The tungsten was brazed to copper for heat dissipation. Since x-rays would be generated in both the tungsten and copper, this gave a degree of freedom to match x-ray beam characteristics for the two targets. The question to be answered with Monte Carlo simulation was whether there was a combination of thicknesses of the two target materials that would result in matched beams over the full energy range.
The Monte Carlo method proved of great help in answering this question. The approach taken was to match angular distributions of energy fluence and spectral distributions on the beam axis for both targets, covering the full range of beam energies. Although angular distributions of energy fluence could be measured relatively easily by measuring dose profiles, measuring spectral distributions is rather difficult (see, e.g., Faddegon et al., 1990). Monte Carlo simulation avoided the need to construct a set of targets with different thicknesses of tungsten and copper and then measure the output and full dose distributions covering the clinical range of field size, including those with beam modifiers (physical wedges, in particular). The simulation details and resulting target design were published (Faddegon et al., 2004). It turned out that a specific thickness of tungsten and copper did yield beam details that were a close enough match to the original target to permit replacement while avoiding the tedious process of recommissioning.

19.3 Target for Imaging

Patients are traditionally set up on the treatment couch prior to irradiation by aligning marks placed on the skin with lasers mounted on the wall and ceiling. The volume in the patient targeted with radiation may move in relation to the surface markers. Modern radiotherapy often involves imaging the patient immediately prior to treatment. A relatively straightforward means to do so is to generate the imaging beam with the same accelerator used to generate the treatment beam. In this way, the radiation therapists setting up the patient can observe the position of the internal anatomy of the patient in relation to the beam portal (Figure 19.3). A CT image may also be obtained with this beam.

The energy range of photons in the x-ray beam that is most suitable for imaging is 10–100 keV. Treatment beams have energies in the range 500 keV through 20 MeV. Lower-energy photons are actually produced in the target used to generate the treatment beam, but these are largely absorbed in the target and never reach the patient. The resulting beam has a reduced surface dose, generally preferred in radiotherapy. A low-atomic-number target may be used to retain these photons for the purpose of imaging.

Monte Carlo simulation has proven instrumental in the design of such a target for planar imaging (Ostapiak et al., 1998) and cone-beam CT (Faddegon et al., 2010). The beam energy is lowered, to fully stop the primary electron beam in the target, as those electrons would otherwise leak through and reach the monitor chamber and even the patient, needlessly adding to the skin dose.

Patient images have been successfully taken with this target for patient alignment with dose in the range 0.3–3 cGy (maximum dose to water at 100 cm SSD for a 10 × 10 cm field given the same monitor units used to image the patient), a low dose compared to the daily treatment dose, usually around 200 cGy. An image obtained with a 4 MV x-ray beam generated in a carbon target for planar imaging (Ostapiak et al., 1998) and cone-beam CT (Faddegon et al., 2010). The beam energy is lowered, to fully stop the primary electron beam in the target, as those electrons would otherwise leak through and reach the monitor chamber and even the patient, needlessly adding to the skin dose.

Patient images have been successfully taken with this target for patient alignment with dose in the range 0.3–3 cGy (maximum dose to water at 100 cm SSD for a 10 × 10 cm field given the same monitor units used to image the patient), a low dose compared to the daily treatment dose, usually around 200 cGy. An image obtained with a 4 MV x-ray beam generated in a carbon target for planar imaging (Ostapiak et al., 1998) and cone-beam CT (Faddegon et al., 2010). The beam energy is lowered, to fully stop the primary electron beam in the target, as those electrons would otherwise leak through and reach the monitor chamber and even the patient, needlessly adding to the skin dose.

Patient images have been successfully taken with this target for patient alignment with dose in the range 0.3–3 cGy (maximum dose to water at 100 cm SSD for a 10 × 10 cm field given the same monitor units used to image the patient), a low dose compared to the daily treatment dose, usually around 200 cGy. An image obtained with a 4 MV x-ray beam generated in a carbon target for planar imaging (Ostapiak et al., 1998) and cone-beam CT (Faddegon et al., 2010). The beam energy is lowered, to fully stop the primary electron beam in the target, as those electrons would otherwise leak through and reach the monitor chamber and even the patient, needlessly adding to the skin dose.

Patient images have been successfully taken with this target for patient alignment with dose in the range 0.3–3 cGy (maximum dose to water at 100 cm SSD for a 10 × 10 cm field given the same monitor units used to image the patient), a low dose compared to the daily treatment dose, usually around 200 cGy. An image obtained with a 4 MV x-ray beam generated in a carbon target for planar imaging (Ostapiak et al., 1998) and cone-beam CT (Faddegon et al., 2010). The beam energy is lowered, to fully stop the primary electron beam in the target, as those electrons would otherwise leak through and reach the monitor chamber and even the patient, needlessly adding to the skin dose.

Patient images have been successfully taken with this target for patient alignment with dose in the range 0.3–3 cGy (maximum dose to water at 100 cm SSD for a 10 × 10 cm field given the same monitor units used to image the patient), a low dose compared to the daily treatment dose, usually around 200 cGy. An image obtained with a 4 MV x-ray beam generated in a carbon target for planar imaging (Ostapiak et al., 1998) and cone-beam CT (Faddegon et al., 2010). The beam energy is lowered, to fully stop the primary electron beam in the target, as those electrons would otherwise leak through and reach the monitor chamber and even the patient, needlessly adding to the skin dose.

Patient images have been successfully taken with this target for patient alignment with dose in the range 0.3–3 cGy (maximum dose to water at 100 cm SSD for a 10 × 10 cm field given the same monitor units used to image the patient), a low dose compared to the daily treatment dose, usually around 200 cGy. An image obtained with a 4 MV x-ray beam generated in a carbon target for planar imaging (Ostapiak et al., 1998) and cone-beam CT (Faddegon et al., 2010). The beam energy is lowered, to fully stop the primary electron beam in the target, as those electrons would otherwise leak through and reach the monitor chamber and even the patient, needlessly adding to the skin dose.

Patient images have been successfully taken with this target for patient alignment with dose in the range 0.3–3 cGy (maximum dose to water at 100 cm SSD for a 10 × 10 cm field given the same monitor units used to image the patient), a low dose compared to the daily treatment dose, usually around 200 cGy. An image obtained with a 4 MV x-ray beam generated in a carbon target for planar imaging (Ostapiak et al., 1998) and cone-beam CT (Faddegon et al., 2010). The beam energy is lowered, to fully stop the primary electron beam in the target, as those electrons would otherwise leak through and reach the monitor chamber and even the patient, needlessly adding to the skin dose.
target is shown in Figure 19.4. Two of the three seeds implanted in the prostate to aid in patient positioning are visible on the slice shown. Note that the metallic implant in the left hip results in minimal artifact in this image, a strength of imaging with MV x-ray beams.

19.4 Electron Beam Magnetic Shield

Many modern linacs are equipped with a powerful magnet at the end of the beam line to bend the accelerated electron beam 270°. Thus, the beam emerges from the evacuated beam line at a right angle to the axis of the waveguide. The bending magnet current is very stable and is set to pass accelerated electrons over a narrow range of energies. For the major Siemens linac models, the resulting full widths at half maxima of the energy distribution peaks are in the range of 6–14% of the beam energy (Faddegon et al., 2009). The standard approach for adjusting the beam energy is to monitor the dose rate while adjusting the RF power in the waveguide and maintaining a constant beam pulse frequency. Given the correct bending magnet current, the RF power that gives the peak dose rate also gives the correct beam energy.

At least for one vendor’s design, the bending magnet field is known to encroach on the path of the beam between the exit window and monitor chamber. This has no effect on x-rays, since they are uncharged and are not deflected by a magnetic field. However, the fringe magnetic field does deflect electron beams. This field must be included in the simulation to accurately match the known asymmetry in the fluence and dose distributions of electron fields. A study of this effect required modification of the BEAMnrc user code to incorporate asymmetry and spatially varying magnetic fields (Shea et al., 2011).

The knowledge that this fringe magnetic field fully accounted for the asymmetry leads to the possibility that a magnetic field could be shielded to produce symmetric beams. A hollow cone machined from ferromagnetic stainless steel (Figure 19.5) was inserted into the treatment head with the linac operated in electron mode. Dose distributions measured once this magnetic field shield was in place were symmetric, as predicted with the Monte Carlo method.

19.5 X-Ray Standard flattener and Extended Flattener

The x-ray beam in radiotherapy is produced by highly relativistic electrons moving within 1% of the speed of light. By conservation of momentum, most of the photons generated in the target travel in a tight cone about the beam axis. A cone-shaped flattening filter is often inserted in the line of the x-ray beam to cut the fluence in the beam center down to match the fluence at the beam edge. Monte Carlo simulation is an established means to design these flattening filters. In one unique application, the technique was used to develop a beam that can be used to image the patient for setup verification and then treat the patient using the exact same beam (Nishimura, 2005). This would permit image-guided radiotherapy using the same beam for imaging and treatment, simplifying the design of the treatment head. Beams generated in low-atomic-number materials are more useful for imaging, having a high proportion of diagnostic energy photons. These photons are low enough in energy to exhibit high contrast to different biological tissues and organs, owing to dramatic differences in the photoelectric effect. These imaging beams could not be fully flattened with a flattener of the same material as the target, since there was insufficient space in the treatment head to accommodate such a flattener. Nevertheless, at the time the study was done, it was felt that partially flattened beams could prove as effective as fully flattened beams in radiotherapy. This
appears to be the case, as unflattened beams are currently under consideration for treatment of even large target volumes, taking advantage of intensity-modulated radiation therapy (IMRT) and modulated arc techniques.

Monte Carlo simulation was used to determine the trade-off of the compromise beams for imaging and therapy by calculating the response of the flat panel detector used for patient imaging (Figure 19.6) and the spatial distribution of photon fluence at the face of the detector as a function of energy. These were used to relate image quality with field flatness for the different beams. Subsequently, one manufacturer incorporated a graphite target into one of the electron foil slots of a multienergy machine (see above), slowing the impetus to further develop a single-beam, imaging-treatment machine.

In a separate application, the existing flattener of a clinical machine was redesigned with the help of Monte Carlo simulation, to fully flatten the x-ray beam out along the diagonals of the largest field (Faddegon et al., 1999). Later, a beveled brass plate was designed as an add-on, placed in the accessory tray slot normally used for wedges, to accomplish the same thing. The plate, designed with Monte Carlo simulation, was used to extend the flat region of the field to cover the largest x-ray field available on the machine: a 40 × 40 cm field at the machine isocenter. The beveled brass plate did produce a nice, flat beam over the whole field, out to 23 cm along the diagonals (Figure 19.7). This plate has been used for total body irradiation (TBI) of patients seated 3 m from the source, as a backup to TBI in a larger room where patients stand up to 4.5 m from the source.

### 19.6 X-Ray Leakage

Radiation from the linear accelerator target is most intense in the direction along the beam axis, where the region in the patient to be treated is generally positioned for treatment. Still, x-rays emanate in all directions from the target, some outside of the area of the treatment beam. Standard practice is for manufacturers to add shielding, usually lead and/or tungsten, to reduce the dose 1 m from the target in all directions to 0.1% of the in-field dose at that same 1 m distance.

The Artiste linear accelerator has a different treatment head than the earlier Oncor model. The distance from the target to the secondary collimators was increased, in part to accommodate full over-travel of the movable jaws that collimate the beam in the Y-direction. Monte Carlo simulation was used to determine the thickness and material of shielding to incorporate in the treatment head. Dose profiles calculated at the depth of maximum dose in the direction of jaw travel with the jaw positioned at full over-travel are shown in Figure 19.8. The results show the magnitude of the leakage from several sources (transmission and scatter from different parts of the treatment head). Note that the study was for a prototype shielding design and does not represent the commercial product.
Monte Carlo for Treatment Device Design

19.7 Electron Secondary Scattering Foils

Monte Carlo simulation was used to consider what changes in the treatment head were needed to enable precise electron beam collimation with the x-ray multileaf collimator (MLC) (Faddegon et al., 2002). The idea was that by bringing the secondary scattering foil closer to the primary foil, and thus reducing the size of the electron field at the secondary foil, one would obtain a sharper penumbra. The replacement of the air downstream of the secondary foil with the much lower scattering helium atmosphere (at the same pressure as air) was also considered. The secondary foil was positioned at its clinical position, 10.6 cm from the primary foil, and at half that distance, 5.6 cm from the primary foil. Scattering foils were designed for treatment with one of three choices of final collimation: the conventional electron applicator, with the distal surface of the electron applicator at a distance of 95 cm from the nominal source position (close to the primary scattering foil), an add-on electron MLC at 65 cm (eMLC), and the x-ray MLC at 35 cm (xMLC).

Monte Carlo simulation was used to design the secondary scattering foil. The design constraints were

1. Similar penetration depth in water, \( R_{50} \) (0.2 MeV energy increase for He)
2. Similar bremsstrahlung tail (it proved sufficient to use the same primary scattering foils)
3. Similar flatness over 30 cm at isocenter without tertiary collimation (40 \( \times \) 40 cm field size setting) to produce beams with flatness equivalent to the clinical beams with tertiary collimation, as shown in Figure 19.9 for the 12 MeV beam

Dose distributions in water were calculated with the Monte Carlo method for the different atmospheres and foil positions and covering the available energy range (6–21 MeV). The three electron collimators were set to the full range of field sizes. The distance from the collimator to the patient, often referred to as standoff, was the same when comparing the penumbral widths. A standoff of 25 cm was used to allow gantry rotation with the patient in treatment position for x-ray IMRT without retracting the tertiary collimator. The width of the penumbra from the calculated dose distributions was largely independent of leaf position and field size. Representative results are shown in Figure 19.10. The penumbral width from the simulated dose distribution was in reasonable agreement for the two configurations measured, limited to the clinically available foil position of 10.6 cm.

The following conclusions were drawn from these results. Replacement of air with helium sharpened the penumbra for all MLC positions (x-ray MLC, extended electron MLC, and applicator). Halving the distance between the foils was only useful when collimating using the x-ray MLC. With the helium atmosphere, the penumbral width under the x-ray MLC was reduced by 35–40%. Of particular interest, collimation of the electron beam with the x-ray MLC provides a clinically reasonable penumbral width when halving the distance between the scattering foils and using helium downstream of the secondary scattering foil. Further experimental validation would be prudent; however, this would require machining the redesigned secondary foil, placing it 5.6 cm from the primary foil, replacing air with helium in the treatment head, and then measuring profiles for the different beam energies at different standoffs, using the x-ray MLC to collimate the beam.
In radiotherapy, a uniform dose distribution is generally sought throughout the clinical target volume. An x-ray field from a single direction produces a uniform dose across the irradiation field with a dose distribution that falls off at a rate of approximately 4% per centimeter. Diametrically opposed fields, known as a parallel-opposed pair, produce a more uniform distribution with depth. A number of situations arise where a wedged-shaped dose distribution is of value, for example, to reduce the dose to structures distal of the tumor by treating at right angles or to adjust for a slope on the patient surface such as in breast treatment (Figure 19.11). The wedge-shaped dose distribution may be produced by placing a wedge made of a dense material such as iron or lead in the beam. The wedge is ideally placed as far away from the patient as possible, to reduce the proportion of dose to the patient from secondary radiation emanating from the wedge. Monte Carlo simulation is ideal for evaluating wedge design and placement.

The wedges for the Artiste were designed with the help of Monte Carlo simulation. The objective was to match the profile measured at 10 cm depth from wedges on the earlier Oncor model. The wedge on the Artiste was positioned further from the target than on the Oncor model, requiring larger wedges to cover the same field size. The Oncor wedge thickness was used for the preliminary design, using the same wedge thickness, scaled to the increased distance of the wedge from the target. A unique design was used to keep the material in the wedge closer to the target, primarily so that the weight of the wedge could be minimized. Results for the thickest wedges are shown in Figure 19.12. The scatter from the Artiste wedges was different enough from the Oncor wedges that some design iteration was needed to match
Monte Carlo for Treatment Device Design

19.9 Electron Applicator Leakage

Linacs used for radiotherapy often are equipped with scattering foils to provide for treatment with electron beams. Electron beams scatter widely, even in air. Fields are generally collimated close to the patient to sharpen the field edges. Monte Carlo simulation has been used effectively to aid in the design of electron applicators (Janssen et al., 2008). Electrons that hit the applicator can be scattered toward the patient or produce x-rays that can reach the patient. This leakage dose needs to be kept to a low level. The International Electrotechnical Commission document IEC 60621-1 (1998) specifies a limit of 1–2% of the maximum dose in the field, the larger value for beam energies exceeding 10 MeV, with the leakage dose average over the area 4 cm beyond the field edge and out to the area shielded by the primary collimator.

An investigation using Monte Carlo simulation was conducted to design a retrofit to an existing applicator (Sawkey and Faddegon, 2008). The objective was to reduce electron leakage without affecting the field used to treat the patient. The BEAMnrc user code was extended to provide for fluence scoring at the sides of the applicator and to accommodate simulation of the retrofit. A retrofit was designed that indeed reduced leakage (Figure 19.13) without affecting the relative output factors or dose distributions used to treat patients. Thus, this retrofit could be used without recommissioning the six electron beams, a very onerous process.

19.10 Electron Multileaf Collimator

Electrons are charged particles that essentially lose their energy evenly as they penetrate tissue, stopping once they exhaust their energy, with little dose delivered to tissue beyond the electron range. They are widely used in radiotherapy to treat more superficial lesions. They may be used alone or in combination with x-ray therapy, often to boost the dose to the region where there may be residual disease following surgical resection of the tumor, a common approach in treating breast cancer. Electrons, being relatively light compared to the atoms in the tissue, present a challenge for treatment planning as they are scattered widely and as the amount of scattering depends on the material traversed, being very different in air, lung, soft tissue, and bone. Monte Carlo simulation, known to accurately account for electron scatter (Faddegon et al., 2009), provides an accurate means for treatment planning in even the most complex radiotherapy geometries.

The availability of accurate treatment planning with Monte Carlo simulation now extends opportunities for electron therapy. Improved conformation to the tumor with significant reduction in harmful energy deposition to healthy tissue may now be achieved by combining x-ray IMRT with a set of adjacent electron fields of different energies (Ge and Faddegon, 2011). Monte Carlo simulation has been used to aid in the design of the add-on MLC to shape the electron fields for this mixed beam radiotherapy (Hogstrom et al., 2004; Shea et al., 2011).

FIGURE 19.12 Prototype wedges based on geometrical divergence alone are compared to the refined design from successive simulations, adjusted to match the wedge profiles for the wedges on the original machine. The calculations are for 0.5 × 2 × 0.5 cm voxels.

FIGURE 19.13 Dose distribution calculated in air just outside the applicator wall from the isocenter up to 45 cm above the isocenter, without (left) and with (right) additional shielding. The isocenter is situated on the central axis at the bottom of the figure.
19.11 X-Ray Image Detection and Processing

X-rays in both x-ray and electron beams provide useful sources for imaging patients. Imaging is useful for positioning the patient prior to treatment and checking and even adjusting for motion during patient treatment.

Patient images, keeping patient comfort and motion (breathing) in mind, are best taken in seconds rather than minutes. Linac gantry rotation speeds place a lower limit on the time taken for a tomographic CT scan. The time for a single rotation is limited to half a minute or more by regulation (European standard EN 60601-2-1, 1998) and as a consequence of the huge weight they support. Current scanning technologies rely on large fields to cover the region of interest in the patient with fewer rotations of the gantry, using cone-beam CT. The large scatter fractions in kilovoltage cone-beam CT scans reduce image contrast and introduce artifacts, complicating the detection and processing of patient images.

Monte Carlo simulation has proven ideal to characterize the primary beam transmitted through the patient and to characterize the radiation scattered from the patient at the point where that radiation reaches the detector. The technique has provided helpful insight for the design, processing, and interpretation of scatter measurements and for the development of scatter correction algorithms (Maltz et al., 2008a,b).

It is also possible to take images of the patient while being treated with electron beams, using the bremsstrahlung generated in the scattering foils (Aubin et al., 2003; Jarry and Verhaegen, 2005). These x-rays, along with a smaller portion of bremsstrahlung generated in the patient, result in a slow fall-off of dose beyond the electron practical range, as seen in Figure 19.14. The x-rays transmitted through the patient have been used for imaging. The bremsstrahlung dose in the open field, just beyond the practical range, is 0.4% of the maximum dose at 6 MeV, increasing to 1.3% at 12 MeV, and 5% at 21 MeV. Thus, lower-energy electron beams have less bremsstrahlung for imaging. Lower energies also produce a flatter bremsstrahlung intensity profile. Patient images were successfully acquired while undergoing electron therapy on a Primus machine.

19.12 Conclusion

The range of examples shown demonstrates the value of Monte Carlo simulation in treatment device design. A word of caution: Monte Carlo simulation, of great advantage in situations involving complex source and geometry configurations, is necessarily complex and therefore prone to error. Experimental validation should be a general rule. Achieving an accurate match, within the combined experimental uncertainty and statistical precision of the calculation, is gratifying and simplifies the design process, giving confidence in the results for the newly designed device. However, the effort required can dramatically slow the design process. In practice, device modification can be done without the need to obtain accurate matches between measurement and simulation. Knowledge of the match (or mismatch) can be sufficient to redesign the device to a stringent specification, by accounting for this difference as part of the redesign procedure. Experience in simulation is essential to judge whether discrepancies with measured data will unduly affect the geometry and material of the device during the design stage. In any case, once the device is designed, it is prudent to verify the design with measurement. In the end, Monte Carlo simulation can improve the efficiency and cost of the design process by reducing or eliminating the need to machine and make measurements with prototype devices. Indeed, Monte Carlo simulation has become a gold standard for fluence and dose calculation in radiotherapy and is arguably integral to the design process.

References


European standard EN 60601-2-1 medical electrical equipment Part 2-1: Particular requirements for the safety of electron accelerators in the range of 1 MeV to 50 MeV (IEC 60601-2-1:1998).


Monte Carlo for Treatment Device Design


20.1 Introduction

Monte Carlo (MC) simulation is considered as the most accurate method for radiotherapy dose calculations due to its capability of faithfully describing the physical processes and flexibly handling complicated geometries. Since the MC simulation is a stochastic method, a large number of particle histories are simulated to achieve a desired statistical accuracy. Despite the vast advancement in computer architecture and the increase of processor clock speed in recent years, the efficiency of the currently available MC dose engines is still not completely satisfactory for routine clinical applications in radiotherapy. One straightforward way to mitigate this issue is to perform the computation in a parallel fashion by taking advantages of advanced parallel computer architectures. By distributing the total computation load to available computing units, it is conceivable that a significant speedup factor can be achieved. Over the years, there has been a considerable amount of research regarding how to implement various MC dose calculation packages on a variety of computing architectures, including multicore central processing unit (CPU), CPU cluster, graphics processing unit (GPU), cloud computing, and so on. The use of those parallel processing techniques for MC dose calculation has offered an attractive approach toward fast MC dose calculation in clinically realistic environments.

The CPU cluster has been the main platform for parallel dose calculations. Under such a platform, one could simply distribute the total number of particle histories to different computing units, which then perform simulations simultaneously and independently of each other. At the end of the calculation, the dose from all the units is accumulated. Although almost all the available MC dose calculation packages were initially developed on a single CPU platform, transferring them to a CPU cluster is straightforward with the help of parallel processing interfaces such as message passing interface (MPI). Currently, parallel versions of a number of available packages, such as MCNP, EGSnrc, PENELOPE, and dose planning model (DPM), have been successfully developed. The expected efficiency gain from such an approach is close to ideal with regard to the number of computing cores used by the simulation (Deng and Xie, 1999; Tyagi et al., 2004; Sutherland et al., 2007).

Lately, a new technology, cloud computing, has also been employed as the parallel processing platform for MC dose calculation. This new platform offers computing resources allocated in an on-demand fashion from a third party, such as Amazon.* In terms of parallel computing strategy itself, cloud-based MC dose calculation is not much different from CPU cluster-based calculations, as the virtual platform is essentially a CPU cluster from the end user’s point of view. This also makes it straightforward to launch available parallel MC packages developed for CPU clusters on the cloud. As such, many packages such as EGS5 (Wang et al., 2011) have been successfully utilized in a cloud environment for dose calculation. The achieved efficiency improvement is very similar to that on a CPU cluster with a certain amount of overhead due to the initializing of the remote virtual cluster and data transfer (Constantin et al., 2011; Keyes et al., 2011; Pratx and Xing, 2011).

20.2 Graphics Processing Unit:
Hardware and Programming

20.2.1 Historical Development of GPU Programming

GPU is a specialized hardware designed to accelerate the processing of graphics information and to write frame buffer that will ultimately be displayed on the screen. The GPU can be present on a discrete graphics card, embedded directly on the motherboard or, more recently, directly on the die of the CPU.

The first GPUs had a fixed pipeline with no control left for the programmer. The graphics primitive information would be sent to the GPU, processed by the fixed pipeline, and sent to the screen. The ability for programmers to have access to the parts of the hardware resources of the GPU directly, rather than through a fixed pipeline, was introduced with the NV20 from NVIDIA (Santa Clara, California, USA) and the R200 from ATI Technologies (now AMD, Sunnyvale, California, USA) in 2001. The programmers could then write pieces of code that would be executed on the vertex and fragment processors, two of the steps present in the graphics pipeline. A distinction, in hardware, was still present between vertex processors and fragment processors (called shaders in computer graphics) and the programmer had the possibility of programming these two units separately. Scientific computing was possible, but difficult, with this generation of the graphics card. Programming was done within a computer graphics context. The programming of the GPU itself was done through languages such as Cg, GLSL, or HSLS, which had the benefit of being at a higher level than the GPU’s specific assembly language.

The NV80 and the R600, in 2006, were the first GPUs to have a “unified shader architecture.” This new feature meant that the distinction between vertex and fragment shaders was gone, replaced by a single type of processor, more general, capable of handling both tasks. Shortly after the introduction of the unified shader architecture, NVIDIA unveiled the CUDA language and AMD (formerly ATI), its stream computing SDK. These two new application programming interfaces (API) gave access to the GPU directly through the C language, without the need of a computer graphics context such as OpenGL or DirectX. In 2008, the specification of the OpenCL API version 1.0 was unveiled.

20.2.2 Hardware Architecture and Threading Model

In what follows, the OpenCL vocabulary is used and the corresponding CUDA vocabulary is put between parentheses; if no corresponding word is given, the same word is valid for both OpenCL and CUDA.

The GPU follows a single-instruction, multiple-data (SIMD) design. As such, it has many processing elements (CUDA cores), split across a number of computing units (streaming multiprocessor). This architecture allows scalability in the number of

...
computing units per device while preserving the same hardware architecture. High-performance GPUs have more computing units than low-performance parts. The GPU, or the GPU and its graphics card, will henceforth be referred to as the device. The system on which the device is installed or, equivalently, the portion of the computer running the C code, will be referred to as the host.

In an SIMD architecture, all the cores of a computing unit must execute the same instruction. A task to be run on the GPU requires data parallelism. For example, doing a term-by-term vector addition of two vectors is a data-parallel task since the same operation (the addition) is applied over a range of data (the terms of the vector). This parallelism is expressed by defining a number of threads required to accomplish the task. Following our example, this number of threads could be equal to the size of one of the vectors. These threads are then grouped in work groups (thread blocks), which are in turn grouped to form a grid. Finally, one kernel, which is a function that runs on the device, is assigned to be executed by one grid. This is illustrated in Figure 20.1. One work group is assigned to one computing unit and multiple work groups can be assigned to one computing unit. If more work groups are required than the computing units can accommodate, the work groups are serialized and the scheduler on the GPU launches them when another work group finishes.

Several memory levels are present on the GPU and on the graphics card in general. Each computing unit has a set of registers. The number of registers is variable over brands and generations. Also present on the computing unit is a pool of local memory (L1 memory), variable in size across brands and generations. This local memory is shared across the threads of one work group and can be used, for example, to store elements needed by several threads in the work group, reducing the need to load from slower memory levels. On newer devices, there is a device-wide data cache (L2 memory) that serves as a hardware-controlled data cache. Finally, outside the GPU but on the graphics card is a pool of global memory. This memory serves as the link between the system memory (system RAM) and the GPU. It is the only memory level writable from the host. Three different types of memory allocation can be done in global memory: linear memory, arrays, or constant memory. Linear memory is assigned by using the equivalent of malloc on the GPU and can be read or written directly by the processing elements. Arrays are allocated and written from the host, after which they are bound to the texture. Textures offer the benefit of a multidimensional spatial locality cache (1D, 2D, or 3D) as well as hardware linear interpolation between elements of the texture. Textures are read-only from the processing elements. The memory levels have been listed in ascending order of latency and descending order of size.

A major issue with MC radiation transport on the GPU is divergence. This will be dealt with in more details in Section 20.3 but the reason for this issue is presented here. As mentioned, the GPU uses an SIMD approach. This means that the same instruction needs to be executed by concurrent threads. This limitation is not GPU-wide; instead, threads of a given work group are split, on the hardware level, in warps. The warp is therefore the smallest granularity and all threads of a warp must execute the same instruction. However, when writing software, it is almost inevitable that branching will occur and that different threads of a warp will have to execute different instructions. The hardware and compiler handle this in two ways: predication or serialization.

FIGURE 20.1 Illustration of componentwise addition of two vectors A and B. A kernel is executed on the GPU grid, and each thread in it computes the addition of one element.
In predication, all instructions are scheduled for execution. For example, in the following pseudocode:

```plaintext
1  if (a < 5)
2     a + 1;
3  else
4     a + 2;
5  end
```

Lines 2 and 4 would be scheduled for execution for all threads of the warp, even if for some the condition is true and for others it is false. A per-thread predicate value of true or false is set according to the value of the condition for instructions 2 and 4. Only the instruction with a “true” predicate will write its result. In serialization, the warp is split into subwarps, according to their branching condition. Threads with a “true” condition will be placed in one warp while threads with a “false” value are put in another thread. Each warp will be launched, one after the other, to execute the branch and reunite after the branch. In both cases, some computational power is lost: in predication, extra instructions are executed, and in serialization, some processing elements will be idle to handle the subwarps.

### 20.2.3 Programming the GPU

#### 20.2.3.1 Available APIs

Two major APIs are available for general-purpose GPU programming: OpenCL and CUDA. OpenCL is a vendor-neutral API managed by the Khronos group. Vendors are then responsible for implementing the OpenCL API within their drivers. CUDA has been developed by NVIDIA, one of the two main vendors of GPUs. NVIDIA alone defines the API, which is only usable on NVIDIA hardware. The same high-level concepts of work groups, processing elements, grids, and memory levels apply to both approaches with only a difference in the vocabulary used to define the elements. A third API is provided by Microsoft, DirectCompute, through its DirectX platform. It is, however, tied to DirectX and Windows programming.

Both these APIs provide an extension to the C/C++ language and are not themselves a language. This allows existing C programs to be modified so that a critical, and parallel, section of the code can be reprogrammed to run on the GPU, while leaving the rest of the program intact.

GPU programming is also available to other programming languages through, mostly, user-maintained libraries. For example, the CUDA.NET and OpenCL.NET provide GPU computing for the .NET platform from Microsoft. PyCUDA allows Python programs access to CUDA. A compiler by the Portland Group allows FORTRAN programs to define kernels written in FORTRAN to run on NVIDIA GPUs.

#### 20.2.3.2 A Bird’s-Eye View of GPU Programming

The broadness of general-purpose programming on graphics processing units (GPGPU) programs is obviously something

---

**FIGURE 20.2** A sample GPU program that adds two vectors into a third vector.
that cannot be covered in a subsection. However, there are general operations, or steps, that are common to most GPGPU programs. The routine steps are the following:

- Allocate memory on the graphics card and move data from the host to the device
- Define the execution configuration (size of work groups and number of work groups)
- Launch the kernel
- Copy the results from the graphics card to the host

If we apply these steps to our example case of adding two vectors into a third vector, we get the program shown in Figure 20.2.

### 20.3 Monte Carlo Simulation on GPU for Radiotherapy

Owing to the stochastic nature of the MC method, simulating an enormous large number of particles is necessary for achieving a desired level of precision for dose calculation. As a consequence, the computation time is prolonged, making this dose calculation method impractical in many clinical contexts, such as online adaptive radiation therapy. Even in some applications where time is not a critical factor, for example, treatment plan verification, having a fast MC dose calculation engine is still highly desirable for a smooth workflow.

Parallel computation is a straightforward, and in many contexts, effective approach to speed up MC dose calculations. MC simulations are known as “embarrassingly parallel,” because they are readily adaptable for parallel computing without the need to rethink the algorithm or the data structures. Specifically for the dose calculation problem, since each computing unit can work on a portion of the total particle histories without interfering with each other, a roughly linear scalability of computational efficiency with respect to the number of computing units is commonly expected. This is indeed the case for the GPU cluster-based MC simulations. For instance, a dose calculation package, DPM, has been ported onto a GPU cluster and almost a linear speedup has been observed with the number of processors, when up to 32 Intel CPU nodes are used (Tyagi et al., 2004).

Nonetheless, this linear scalability seen on the CPU cluster is hardly achievable on GPU architectures due to an inherent conflict between the special architecture of the GPU and the random nature of MC simulation. In general, the means of performing parallel computation are categorized into task parallelization and data parallelization. MC simulation, a typical task parallelization problem, is preferable for a CPU cluster using, for example, the MPI. For the MPI implementation on a CPU cluster, all the particle histories simulated in an MC dose calculation can be equally divided and distributed to all the processors, which can then be executed simultaneously without interfering with each other. Only at the end of the computation will the dose distribution be collected from all the processors. This method of parallelization is apparently capable of speeding up the simulation easily on a large number of CPUs. On the other hand, a GPU is suitable for the data parallelization approach. As mentioned previously, a GPU multiprocessor employs an architecture called SIMD (NVIDIA, 2009), under which the multiprocessor executes a program in groups of 32 parallel threads termed warps. If the paths for threads within a warp diverge due to, for example, some if–else statements, the warp serially executes one thread at a time while putting all other threads in an idle state. Thus, high computation efficiency is only achieved when 32 threads in a warp proceed together along the same execution path. In the aforementioned example of adding up two vectors, we can simply have each GPU thread add two corresponding entries in the two vectors. This is essentially performing the same operation on different data set, where no thread divergence occurs and high efficiency can be expected. In contrast, in an MC calculation, the computational work paths on different threads are statistically independent. It is quite hard, if not possible, to explicitly control the operations at each thread within the stochastic nature of the MC simulation. This vast divergence between GPU threads, if not treated carefully, will destroy the parallel nature of GPU computation to a large extent, leading to low computational efficiency. A GPU-based MC simulation package, gDPM v1.0, was developed (Jia et al., 2010), where the GPU threads are simply treated, as if they were independent computational units in a CPU cluster. Despite the large number of GPU threads used in the simulation, only 5–6 times speedup was observed, indicating that special care has to be taken when implementing an MC engine on the GPU.

To mitigate this incompatibility between the GPU data-parallel architecture and the MC task-parallel paradigm, one has to seek for new simulation strategies to achieve a high efficiency. While this issue has not yet been completely resolved and is still an active research topic, there are already some effective approaches to partially relieve the problem.

There are two types of GPU thread divergence that one may encounter in an MC simulation for dose calculation. First, owing to the different particle transport physics for different types of particles, for example, between electrons and photons, one has to perform simulation in different manners. The simultaneous simulations of different types of particles will then result in divergence between threads. Second, even between particles of the same type, thread divergence still occurs because of the randomness of the particle transport process. For example, the particle transport trajectory, number of transport steps, and interaction type vary significantly from one thread to another. While we have no explicit control on the divergence of the second type, by programming the simulation carefully, it is possible to mitigate the first one. Recently, Hissoiny et al. (2011a) have developed an MC dose calculation package, GPUMC, where the idea of separating the simulations of electrons and photons was first proposed. By smartly placing the particles to be simulated into two arrays holding electrons and photons separately and having the GPU simulate particles in only one array at a time, a considerable amount of speedup has been reported. In another recently developed
package, gDPM v2.0 (Jia et al., 2011), such a strategy was also employed and was again found effective.

### 20.4 Current GPU-Based Monte Carlo Codes

GPU-based MC is a new area of research in radiotherapy and as such not many codes currently exist. They can be divided into two categories: photon codes and coupled electron–photon MC codes.

#### 20.4.1 Photon Monte Carlo Codes

Photon-only MC codes were the first to appear in the literature due to the relative ease of implementation that such codes involve. They have first appeared in the domain of light transport in biomedical optics.

Alerstam et al. (2008b) have the earliest publication relating to MC light transport on GPU. Their work consisted in porting an existing CPU photon migration code to the GPU (Alerstam et al., 2008a). Briefly, photons are launched toward a medium and tracked until the photon escapes the medium or its total time of flight exceeds a user-defined maximum value. In their short letter, they report accelerations of up to 1080× when comparing to a CPU implementation of the same code. This project has later evolved in the CUDAMCML project, which is a GPU implementation of the Monte Carlo modeling light transport (MCML) photon migration code (Wang et al., 1995). A comparison between the CPU and GPU code yielded up to a 272× acceleration. At that point, CUDAMCML could only handle layered geometries with a maximum of 100 layers, stored in constant memory. The manual of the platform details the first attempt at reducing divergence in GPU MC applications by rethinking the overall problem. Indeed, the implementation used on the GPU deviates from the original MCML implementation’s handling of layer crossing, which would not have been efficient on the GPU.

This project again evolved in GPU–MCML (Alerstam et al., 2010). In this version of the work, the authors incorporate the previous developments in light transport (Badal and Badano, 2009; Fang and Boas, 2009; Lo et al., 2009; Ren et al., 2010) and add new developments to reduce divergence and other bottlenecks. One such development is the storing of the portion of the dose grid near the photon source in shared memory. If a photon is absorbed in this region, its contribution is summed in shared memory. With GPU–MCML, they observe an acceleration of up to 869× compared to a CPU version. The shared memory optimization alone increased the acceleration factor from 260× to 620×. Their paper also presents the first multi-GPU MC code implementation. Close to the ideal linear gains with regard to the number of GPUs used are reported.

Finally, GPU–MCML was modified to accept complex three-dimensional (3D) geometries (Lo et al., 2009). The geometry array is however still located in constant memory, which reduces the size of the geometry modeled due to the relatively small size of the constant memory. The authors show an acceleration of 270× compared to a conversion of the GPU program to a serial CPU program.

An independent work by Fang and Boas (2009) was also targeted at light transport in a 3D environment. Their solution departs from storing the geometry information in constant memory and instead places it in global memory, incurring the performance penalty associated with the lower bandwidth and higher latency. The authors also present a study on the error rate of using atomic versus nonatomic writes to global memory to update the scoring grid. They show that at a distance further than three voxels away from the source, the error induced by using nonatomic operations is around 1%. The authors show a speedup of 100× through 250× for their version not using atomic writes.

Badal and Badano have ported the photon transport part of PENELope (Salvat et al., 2009) to the GPU (Badal and Badano, 2009), yielding MCGPU code. The implementation is targeted to radiographic image creation. They make use of the Woodcock tracking algorithm (Woodcock et al., 1965). Their test simulation consists of launching 50 keV photons at a 3D phantom and collection results in a two-dimensional (2D) detector. They report a speedup of 27× compared to a modified version of the PENELope code using the Woodcock algorithm. The code is made freely available by the authors.

Recently, Jia et al. (2012b) have developed photon transport codes gCTD, which targets at the estimation of the imaging radiation dose to a patient from computed tomography (CT) or cone beam computed tomography (CBCT) scans. As opposed to MC GPU that translates PENELope directly into CUDA, the development of gCTD is optimized for GPU programming. A number of sampling algorithms are designed to yield a high efficiency. Moreover, gCTD supports the simulation of CT/CBCT scanners to a high level of realism, including the modeling of source spectrum, fluence map, and so on. Since no corresponding CPU codes are available, the speedup factors are measured against EGSnrc, a commonly used CPU MC simulation package, where electron transport in EGSnrc is turned off for a fair comparison. It was observed that gCTD is about 76 times faster than EGSnrc in a realistic patient phantom case. As for the absolute computation time, the radiation dose to a patient in a CBCT scan can be computed in ~17 s with less than 1% relative uncertainty.

Another package gDRR is also developed by Jia et al. (2012a) for accurate and efficient computations of x-ray projection images in CBCT under clinically realistic conditions. Simulation of x-ray projection images plays an important role in CBCT-related research projects, such as the design of reconstruction algorithms or scanners. gDRR accurately computes the three key components in a CBCT projection image, namely, primary, scatter, and noise signals. In particular, a GPU-based MC simulation is utilized to obtain the scatter signal and to facilitate the calculations of the noise signal. gDRR uses the same MC photon transport functions as in gCTD, with some necessary modifications. It also supports very flexible scanner geometry as well as some details of the scanner, such as a finite x-ray focal spot size.
detector response, and so on. For a typical CBCT projection with a polyenergetic spectrum, the MC simulations take 28.1–95.3 s, depending on the voxel size. The MC-simulated scatter signal using gDRR is in agreement with EGSnrc results with a relative difference of 3.8%.

Finally, Hissoiny et al. have developed bGPUMCD, a dedicated GPU package aimed at brachytherapy dose calculations. It is useable for both low-dose-rate (Hissoiny et al., 2011b) and high-dose-rate brachytherapy dose calculations. For LDR calculations, a mixed simulation with both analytical geometrical description of the seeds and voxelized patients is used. TG-43 parameters have been reproduced by GPUMCD and compared to values found in the literature and in the TG-43 report. GPUMCD was found to be able to reproduce radial functions within 1.25% and anisotropy functions within 4% for low-energy single-seed simulations. Phase spaces can also be used to remove the need to track particles through their seed. For HDR dose calculations, bGPUMCD uses a phase space to represent the seed in a purely voxelized geometry. Single dwell position dose kernels as well as full dose calculations using preoptimized plans and real patient geometries have been performed. bGPUMCD is able to compute single dwell position dose kernels within 0.5 s and full dose calculations within 2 s for a 2% statistical uncertainty. Dosimetric parameters, such as D90 and V100, are within 2% when compared to a reference GEANT4 dose calculation.

20.4.2 Coupled Electron–Photon Monte Carlo Codes

Three GPU-coupled electron–photon MC codes have been developed and targeted at radiotherapy applications: GPUMCD (Hissoiny et al., 2011a), gDPM (Jia et al., 2010, 2011), and GMC (Jahnke et al., 2012).

20.4.2.1 GPUMCD

GPUMCD is a new implementation of the MC simulation based on existing physics. The code, the overall algorithm, the inputs/outputs, the memory management, and others have all been tailored specifically for the GPU while the sampling routines come from existing general-purpose MC packages such as EGSnrc (Kawrakow, 2000), EG55 (Hirayama et al., 2010), and PENELOPes (Salvat et al., 2009).

The photon simulation is done in an analog manner. GPUMCD models Compton scattering, the photoelectric effect, and pair production. Compton scattering uses a free atom approximation and the Klein–Nishina cross section. The sampling routine is derived from the method by Everett et al. (Everett and Cashwell, 1971). The photoelectric effect is modeled simplistically by assuming that every electron is ejected from the K-shell and by ignoring atomic relaxation and fluorescence. The angular sampling of the electron is selected according to the Sauter distribution (Sauter, 1931) following the sampling formula presented in Section 2.2 of the PENELOPe manual (Salvat et al., 2009). Pair production events are modeled by ignoring the production of positrons. Instead, two electrons, with randomly selected energies within the incoming photon energy, are generated. The angle of both electrons is sampled using the algorithm presented in Equation 2.1.18 of the EGSnrc manual.

The electron simulation is performed using a class II condensed history approach (Berger, 1963). The so-called hard interactions are simulated explicitly while soft, subthreshold, interactions are modeled in a condensed manner. Inelastic collisions are modeled using the Møller cross section, and the sampling routine presented in Section 2.4.3i of the EGSnrc manual (Kawrakow et al., 2011) is used to sample the energies and orientations of the electrons. The photon created during a bremsstrahlung interaction has its energy sampled following the sampling routine presented in Section 2.4.2.ii of the EGSnrc manual (Kawrakow et al., 2011). The EGS4 approximation to the photon angle is used in which the photon angle is assigned as \( \theta = m_e c^2/E \), where \( E \) is the incoming electron kinetic energy.

Multiple scattering is modeled by using the methods of Kawrakow and Bielajew (1998) to select a scattering angle and PENELOPes to perform the electron step. The random hinge method is used to transport the particle: a given electron step is randomly split in two parts; the electron advances to the hinge point, is deflected, and advances for the rest of the step length. At the hinge point, the angle selected using the method of Kawrakow and Bielajew is used to deflect the electron.

GPUMCD can handle voxelized or parametric geometries as well as a mix of the two (Hissoiny et al., 2011b). Parametric geometries include secondary quadratic surfaces as well as boxes. GPUMCD also supports a large number of particle sources (monoenergetic, polyenergetic, phase space based, parallel, divergent, sweeping beamlets, multiple beams, arbitrary orientations, etc.) and the C++ approach to particle source definition makes it easy to include a new source type.

The GPUMCD implementation of the MC algorithm is different compared to the other existing MC packages since it had to be redesigned to fit the GPU architecture. The major difference is in the way particle histories are sequenced. In most platforms, a primary particle is launched and its secondary particles are put on top of the particle stack, regardless of their type (electron, photon, or positron). An electron track simulation could then be immediately followed by a photon simulation, and so on. On the GPU architecture, many particles are simulated at once, where the number of particles in parallel depends on the number of processing elements of the GPU. Also, since the GPU is an SIMD machine and all threads of a warp must execute the same instruction flow, simulating electrons and photons at the same time would lead to a high degree of divergence. Similarly, if all secondary particles of a primary particle are simulated on the spot, a large degree of divergence in the number of secondary particles per primary particle could be found. These two factors would lead to a large degree of performance loss on the GPU. For these reasons, GPUMCD implements a dual stack of particles: one for electrons and one for photons. Every secondary particle created during a particle track simulation is placed on its corresponding stack for a later simulation. This way, each thread only simulates one particle. Also, particle stacks are treated individually, which
ensures that at a given time, only electrons or only photons are being simulated. The overall workflow of GPUMCD is shown in Figure 20.3. The size of these stacks is graphics card dependent and varies with the amount of global memory available on the graphics card. For example, a 1.5 Gb pool of global memory can accommodate $8 \times 2^{20}$ particles per stack while leaving enough memory for the other elements required by the simulation. Such elements include cross-section data, stored as one-dimensional (1D) textures, phantoms of material identifiers and densities per voxel, stored as 3D textures, pseudorandom number generator states, stored in registers and global memory, and a large number of simulation parameters, stored in constant memory.

GPUMCD has been benchmarked in a number of situations. In voxelized slab geometries with simplistic monoenergetic sources (Hissoiny et al., 2011a), it has been found to be within a 2%–2 mm gamma criteria of EGSnrc in 98% of all voxels or more for most cases. On a per particle basis, it has been benchmarked to be around 1000 times faster than EGSnrc for photons simulation and 1500 times faster for electrons simulations. Compared to the DPM platform (Sempau et al., 2000), its speed is around 400 times faster for photons simulations and 250 times faster for electrons simulations.

In voxelized patient geometries (Hissoiny et al., 2011c), it is able to compute a dose distribution within a 2% statistical uncertainty in less than 8 s per beam or 11 s per plan. GPUMCD has also been adapted to be the dose calculation tool of the magnetic resonance imaging (MRI)-linac by incorporating the effects of the magnetic field on charged particles (Hissoiny et al., 2011c). Within this dose calculation scenario, it is able to compute the dose distribution in less than 15 s per plan within a 1.5T magnetic field.

**20.4.2.2 gDPM**

gDPM, another GPU-based MC package for coupled photon–electron transport, was developed by Jia et al. (2010, 2011). Such a package tries to port a publically available CPU-based fast MC package, DPM, to the GPU platform. In this process, DPM particle transport physics is maintained and hence the simulation accuracy, whereas the code is optimized for the GPU architecture to improve computational efficiency.

The original sequential DPM MC code was previously developed for fast dose calculations in radiotherapy treatment planning (Sempau et al., 2000). It aims to simulate coupled photon–electron transport with a set of approximations valid for the energy range considered in radiotherapy. Specifically, the photon transport is handled by using Woodcock tracking method, which greatly increases the simulation efficiency of the boundary tracking process (Woodcock et al., 1965). As for the electron transport, DPM implements a condensed history technique. Step-by-step simulation is used for inelastic collisions and bremsstrahlung emission involving energy losses above certain cutoffs. It also employs new electron transport mechanics and multiple scattering distribution functions to allow long transport steps. Continuous slowing down approximation is employed for energy losses below some preset energy thresholds. Positron transport is treated in the same way as for electrons and two photons are created at the end of the positron path to account for the annihilation process. The accuracy of DPM has been demonstrated to be within ±2% of measurements for both clinical photon and electron beams (Chetty et al., 2002, 2003).

The first version of gDPM was developed in 2009 (Jia et al., 2010). In this version, all GPU threads are treated as if they were independent computational units. All the source particles are distributed to the GPU threads and each thread tracks the entire history of a source particle as well as all the secondary particles it generates. It is found that only 5.0~6.6 times speedup can be achieved due to the intrinsic conflicts between the randomness of the MC simulation and the GPU SIMT architecture mentioned previously.

Lately, the gDPM v2.0 was released with a higher computational efficiency (Jia et al., 2011). gDPM executes the simulations in a batch fashion, where the total number of particle histories are divided evenly into a number of batches. The simulations are conducted for each batch sequentially and the dose depositions are recorded. Statistical analysis is performed based on the results in all batches to obtain the average dose to each voxel and the corresponding uncertainties. The overall workflow is shown in Figure 20.4. Within each batch, the simulation details are presented by the workflow in the right panel of Figure 20.4. Since photons and electrons undergo different physics and hence attain different
execution paths, gDPM v2.0 uses a simulation scheme where photon transport and electron transport are separated to partially relieve the thread divergence issue, as first proposed by Hissiony et al. Specifically, a particle array of length $N$ is allocated to store all the particles currently being simulated. The particles in this array can be either photons or electrons, but all of the same type at any time during the simulation. The size of this particle array should be large enough, so that the GPU can fully exploit its parallelization ability, while not too large to fit in the GPU memory. Moreover, since the GPU executes simulation in warps, that is, a group of 32 threads run simultaneously on a multiprocessor, it is beneficial to choose $N$ to be a multiple of warp size to avoid wasting resources, for example, $N = 65,536$ in a typical setup. Meanwhile, two more arrays are allocated as stacks to store any secondary particles generated during the simulation, one for photons and one for electrons. Their lengths are large enough to hold all secondary particles generated during the simulation, one for photons and one for electrons. Their lengths are large enough to hold all secondary particles. During simulation, a total number of $N$ GPU threads are launched simultaneously to transport $N$ particles of the same type, while putting all secondary particles into the corresponding stacks, which effectively separates the simulations of the photon and the electron transport. Besides this simulation strategy, a high-performance random number generator supported by NVIDIA (NVIDIA, 2010) and hardware supported linear interpolation are also utilized to further speed up calculations.

The development of gDPM v2.0 also emphasizes on its clinical practicality by integrating various key components necessary for dose calculation in radiotherapy. An interface of the gDPM package is built to load clinical IMRT or VMAT treatment plans in the format of DICOM RT. This includes the functions of loading voxelized patient CT data, organ structure information, fluence map of a treatment plan, and so on. Computations related to the geometry of a treatment plan are enabled to take into account the rotations of linac gantry, multileaf collimator (MLC), collimator, and couch.

As for source modeling, two simple aspects have been considered in gDPM in the current released version. First, it is supported to generate source particles according to a designed photon fluence map in IMRT or VMAT plans. Such a function is achieved by using Metropolis sampling algorithm (Hastings, 1970), which is favored in GPU due to its infrequent memory access. Second, photon source particles are generated according to a given energy spectrum. Specifically, the total number of photons is distributed to a few energy intervals according to the spectrum and simulation is performed for each interval sequentially. This strategy ensures that, at any moment of the simulation, all GPU threads are dealing with particles of similar energies, removing the possibility of losing efficiency due to the variance of simulation time between GPU threads handling photons of different initial energies.

Yet, accurate source modeling of a linear accelerator is not a trivial problem. gDPM v2.0 leaves the interface of source particle generation part open and users can supplement their own functions to generate source particles according to their own source model or simply by using a phase space file.

As for the validation of gDPM, simulations in various phantom cases indicate that results from CPU DPM and gDPM are in
good agreement for both the electron and the photon sources. In particular, statistical \( t \)-tests are performed and the dose differences between the CPU and the GPU results are found not statistically significant in over 96% of the high-dose region and over 97% of the entire phantom region. Speedup factors of 69.1–87.2 have been observed against a 2.27 GHz Intel Xeon CPU processor. An IMRT or a VMAT plan dose calculation using gDPM can be achieved in 36.1–39.6 s with a single GPU for the average standard deviation less than 1% of the prescription dose. Moreover, multi-GPU implementation of gDPM has also been developed. Another speedup factor of 3.98–3.99 compared to a single GPU has been observed using a four-GPU system.

### 20.4.2.3 GMC

Recently, another MC simulation package, GMC, has been developed (Jahnke et al., 2012). It is a GPU implementation of the electromagnetic part of the Geant4 MC code (Agostinelli et al., 2003). The main difference between this code and the aforementioned two packages is the GPU implementation of lepton (electron and positron in this code) transport. A full history of a lepton consists of a large number of small steps separated by voxel boundaries or discrete interaction sites. As opposed to having a GPU kernel simulate, a full history of a lepton as in GPUMCD and gDPM, a GPU kernel in GMC loads leptons from a stack in global memory, transports each of them by only one step, and writes them back to the stack. Such a kernel is repeatedly invoked to move the leptons forward till they are absorbed or are outside the phantom region. The advantages of this simulation scheme are the reduction of thread divergence due to different history lengths among particles. In particular, if a GPU thread handles an entire particle history, those threads with short-lived particles will wait for others, resulting in reduced efficiency.

Another unique feature of GMC is random number generation. GMC utilizes a modified Mersenne twister algorithm for pseudorandom number generation due to its very long period. The random numbers are generated by a dedicated kernel and are not on the fly. As such, a number of random numbers are pregenerated and stored in a pool residing in the global memory. Other particle transport kernels read these numbers whenever necessary. Once those pregenerated random numbers are exhausted, a GPU kernel is invoked to refresh the random number pool. It was reported that a 3–4 times of speedup in terms of total computation time has been achieved by this method compared to the on-the-fly generation of random numbers. This can be attributed to the fact that removing random number generations in particle transport simulations reduces the degree of GPU thread divergence.

GMC has been validated in a number of contexts ranging from a simple homogeneous water phantom with an open beam to a real patient IMRT case with a high degree of fluence modulation. A phase space source model is used in those testing cases, which is generated by simulating an Elekta Synergy linac head using Geant4 on the CPU. In all the phantom cases, the agreement between the GMC results and the Geant4 results is satisfactory with over 97.5% region passing the 2%/2 mm gamma index test. The computational speed is assessed by executing GMC on an NVIDIA GeForce GTX580 card. On an average, GMC processes 657.60 histories per millisecond. Compared with the GPU execution of Geant4 on a 2.13 GHz Intel Core2 processor, a speedup factor of 4860 is achieved. This large speedup factor can be partly ascribed to the slow Geant4 simulations on the CPU. Furthermore, a linear speedup with respect to the number of GPUs is also observed.

### 20.5 Summary

In this chapter, we have discussed the applications of GPU for fast MC dose calculations in cancer radiotherapy. Some basic principles of the GPU architecture and GPU programming have been introduced, as well as those difficulties when applying GPU for MC simulation. It has been seen that the GPU architecture is not ideal for conducting MC simulations in radiotherapy. Developing GPU-friendly packages for MC simulations therefore requires a redesign of the simulation algorithms. A number of GPU-based MC packages developed for a variety of purposes in radiotherapy have become available. The achieved high performance with affordable GPU cards has undoubtedly demonstrated the unique advantages of the GPU in radiotherapy.

At the end of this chapter, we would like to present a few directions to be explored in the near future for GPU-based MC simulations in radiotherapy. Generally speaking, there are two aspects one needs to keep in mind to improve the practicality of a GPU-based MC simulation package in clinical environments: efficiency and accuracy. Those future directions essentially target the improvements of those two aspects.

First, it is highly desirable to further boost the simulation efficiency. Despite the accomplishments achieved so far, MC simulation can only preliminarily satisfy the time requirements in radiotherapy for the purposes of dose calculation and treatment verifications. In many other applications where repeated dose calculations are needed, such as computing dose deposition coefficients for IMRT planning, the current GPU-based MC simulation is still far away from being clinically applicable. There are a number of ways for potentially increasing the MC simulation efficiency besides those trivial approaches of using more and better GPUs. In particular, it is, and will continue to be, an active research topic to design GPU-friendly simulation schemes to further alleviate the branching issue and to use the GPU’s memory more efficiently. Also included in this direction are utilizing variation reduction techniques, and designing new ones suitable for GPU if needed, to increase the simulation convergence rate and hence the overall efficiency. Those variation reduction algorithms might be of particular use on GPU, as they usually tend to force the simulation be performed in a more predictable fashion and therefore may encounter less branching problems among different GPU threads.

Another direction is performing simulation with a higher accuracy in the underlying physics to obtain better simulation results. Previously, constrained by the limited computing power, a number of components in MC simulations were simplified or even
ignored. Armed with the powerful GPUs, it becomes possible for us to retrieve those components, if needed, to yield simulation processes that are closer to the realistic cases and hence potentially more accurate results. For example, the simulation accuracy is largely dependent on how realistically the source particles are generated. In many current MC packages, it is popular to empirically model the linear accelerator head, that is, energy spectrum, fluence distribution, angular distribution, and so on, and generate source particles accordingly. Detailed simulations of MLC transmission using particles from a phase space file obtained from a separate MC simulation are another possible but complicated approach. With the computing power of GPU, we can potentially perform the particle transport starting from the phase space file above the MLC, while considering many effects such as MLC transmission, interleaf leakage, MLC round end leakage, and so on. It is expected that those components will greatly enhance the simulation accuracy and hence facilitate the clinical applications. Yet, it is a legitimate question regarding to what extent of realism one needs to perform MC simulations. It is highly likely that the improvements of accuracy from many additional detailed simulations are incremental, especially considering there exist other error sources on the simulation results, such as statistical uncertainties from the MC method itself. Nonetheless, the GPU at least offers us a means to consider those components that are necessary but was simplified or ignored previously. It is an interesting research topic to identify those crucial components responsible for the accuracy and implement them in the simulation.

References


21.1 Introduction

Shielding is an important and challenging aspect of radiotherapy management. The purpose of shielding radiotherapy machines is to reduce the dose equivalent to a point outside the treatment room to an acceptably low level. Failure to design shielding adequately can result in overexposure to personnel or expensive barrier retrofitting.

The methods for calculating treatment vault shielding for conventional radiotherapy accelerators are documented in the National Council on Radiation Protection and Measurements (NCRP) reports (NCRP Report Nos. 51, 79, and 151). The approach relies on transmission, leakage, scatter, and albedo factors based on standard room layouts and designs. The transmission factors are based solely on ordinary concrete (NCRP Report No. 151) with a density of 2.35 g/cm³ (147 lb/ft³), and have been generated for radiotherapy beams of only a few nominal energies. Between these and other approximations detailed below, they can only be considered as approximations of the true shielding requirements of a given medical accelerator. Different compositions of concrete or other shielding materials are often used, and different accelerators have notably different spectra, even at the same nominal energy (Sheikh-Bagheri and Rogers, 2002). Despite these reservations, the calculations have been shown to be quite conservative in practice (IAEA, Low 2003). However, there are many situations in radiotherapy shielding that cannot be calculated with NCRP formalisms, or where it is of interest to determine shielding requirements more precisely. In addition, there are now many nonstandard accelerators (e.g., Tomotherapy®, Cyberknife®, and Gamma knife®) as well as nonstandard beams (e.g., flattening filter free (FFF)). For example, Kry et al. (2009a) have shown, via Monte Carlo, substantial differences in 10th-value layers (TVLs) between flattened and FFF beams. There are also nonstandard geometry rooms for which the standard NCRP formalisms do not easily apply, for example, a primary beam striking the outer maze wall. Additionally, while the NCRP offers guidance on proton and light-ion particle therapy vault shielding (NCRP Report Nos. 51 and 144), the novelty of such equipment makes shielding studies of considerable current interest. For these types of scenarios, the formalisms of the NCRP reports may be insufficient and Monte Carlo methods are a viable option to examine the shielding requirements of a given situation.

There are two approaches to solving radiation shielding problems using the Monte Carlo method. First, Monte Carlo may be directly applied to the solution of a shielding problem by modeling the specific accelerator, vault, and potentially even the patient or phantom of interest. This approach offers the most accurate solution for a given problem by accounting for the specifics of the radiation beam and environment. However, attaining sufficient details of all aspects of the radiotherapy system and its environment is extremely challenging. Sufficient details should be readily achievable for the primary and scattered radiation; however, for simulations of head leakage, this can be very demanding. As head leakage is often the dominant factor for secondary barriers, particularly for intensity modulated radiation therapy (IMRT)
treatments, and is important for direct door shielding, its accurate calculation is essential. Monte Carlo models that might be sufficiently accurate to calculate head leakage levels are complex and therefore uncommon.

The second approach using Monte Carlo is to solve for general shielding parameters and from these derive an analytic solution that allows for the shielding requirements of particular cases to be solved. This is the approach presented in NCRP Report No. 151, where the Monte Carlo method has been used to generate shielding parameters, such as patient scatter fractions, for a general situation. This second approach has the advantages of requiring a less complex model and being less computationally intense. However, this is not a general solution and only solves part of the overall problem.

This chapter reviews the rationale and special considerations for performing shielding-relevant calculations using the Monte Carlo technique. Section 21.2 describes the Monte Carlo techniques as applied to shielding, Section 21.3 details photon shielding calculations, Section 21.4 deals with neutron calculations, Section 21.5 addresses applications of Monte Carlo for nonstandard room geometries, Section 21.6 addresses the application of the Monte Carlo approach for nonstandard beams and accelerators, including protons and light-ion accelerators, and Section 21.7 provides a summary of this chapter and future considerations for Monte Carlo research in therapy shielding. Finally, while considerable Monte Carlo work has been carried out for radiation shielding in the field of nuclear reactors (e.g., Jaeger et al., 1975), this chapter concerns itself only with the work that is specific to radiotherapy installations.

21.2 Monte Carlo Codes

21.2.1 Specific Programs

A wide range of different Monte Carlo codes have been used for shielding calculations in radiotherapy, including EGS/EGSnrc (Kawrakov et al., 2010), MCNP/MCNPX (Briesmeister 1993), GEANT4 (Agostinelli et al., 2003), ITS (Halbleib and Mehlhorn 1998), and others. Typically, any of these codes is a viable option. Naturally, different codes have different strengths and weaknesses and some considerations are presented below.

Computation time is an important consideration. For primary barrier attenuation calculations, calculation of the third or fourth TVLs may require the simulation of considerably more particles than would be typically needed for patient dose calculations. Calculations involving secondary barriers may be even more time consuming as they will typically involve relatively low-frequency scattering events or head leakage and only a tiny minority of generated and tracked particles may even impinge on the barrier. Codes that handle radiation transport most expeditiously are clearly advantageous. Moreover, variance reduction techniques will typically be an important part of such simulations.

The flexibility of the geometry specification is also a potentially important consideration. For many shielding applications (Sections 21.3.1 through 21.3.3), the barrier transmission can be calculated accurately using a detailed simulation of the radiation transported through the beam line accelerator components. Therefore, Monte Carlo systems that readily model beam line components hold an advantage (e.g., BEAMnrc). However, if head leakage or neutrons are to be simulated, modeling of non beam line components of the accelerator head (primary collimator, structural and shielding components, etc.) may also be required. In this case, a Monte Carlo code with more flexible geometry modeling (e.g., MCNP/MCNPX) is likely to be much more useful. Similarly, if arbitrary vault designs are being considered, flexible geometry specifications are again important. The need for a flexible geometry definition clearly depends on the specifics of the shielding problem being evaluated.

Finally, the ability of the code to track a variety of particles may be an important consideration. Codes including MCNP/MCNPX or GEANT can track a wide range of particles (although not all materials may have the available cross sections, e.g., photoneutron production in steel may require additional cross-section data in MCNPX), whereas EGSnrc only tracks photons, positrons, and electrons (thereby neglecting neutrons). In general, only some cross sections will be tabulated in any cross-section library, while other cross sections may be approximated with models. Users of any code should be aware of the limitations in the physics employed in their Monte Carlo system.

21.2.2 Interpretation of Results and Uncertainties

In any Monte Carlo simulation, calculated tally values have an associated uncertainty. This uncertainty will be a combination of both random errors (which are readily estimated by most MC codes and typically quoted) and systematic errors (that reflect uncertainties in the modeling geometry, source specification, cross-section data, etc.). In shielding problems, the random/statistical uncertainty may be minimized with large numbers of histories and the use of variance reduction techniques. In that case, systematic errors may dominate and an estimate of the systematic uncertainty should be incorporated into the uncertainty of the final result.

21.3 Photon Shielding

Shielding of photons must include primary photons, as well as those scattering off a patient or wall onto a secondary barrier (patient scatter and wall scatter, respectively), and leakage from the accelerator head. Both the intensity of the photons and their energy (TVL) affect the shielding burden of each of these sources of radiation. In the following sections, these two approaches to shielding calculations (full Monte Carlo simulation of the linear accelerator and vault or piecemeal determination of individual shielding parameters) will be discussed and compared.
21.3.1 Barriers

For radiotherapy shielding, the most common shielding material is concrete. Concrete is a mixture of cement, water, and aggregate. The quantity and type of aggregate can vary yielding many different types of concrete. Even concrete of the same name can show substantial variation: “ordinary” concrete can vary in density between 2.09 and 2.50 g/cm³ (NCRP Report No. 151). Additionally, high-density aggregates (including iron) can be used to make heavy concretes that offer increased photon attenuation. The atomic composition of several types of concrete (by mass fraction) is shown in Table 21.1. Similar tables are available elsewhere (see, e.g., Kase et al., 2003 or Jaeger 1975).

While these tables allow for simple calculations, the reality of a particular concrete wall is invariably more complicated as even nonuniformity in the distribution of aggregate can notably affect the shielding capabilities of concrete (Kase et al., 2003). While this poses challenges for the application of Monte Carlo to concrete shielding, this is also a challenge for analytical shielding approaches. Monte Carlo approaches have a distinct advantage in the determination of TVLs or half-value layers (HVLs) of high-density concretes. The attenuation (and therefore the amount of shielding material) of such concretes cannot simply be scaled by the density of the medium. Because of the high-Z components in high-density concrete, the true TVL of high-density concrete is less than that predicted by density scaling the TVL of low-density concrete. This was noted by Facure et al. (2007) who found the TVL to be overestimated by up to 20% when the low-density concrete TVL was scaled by the relative density.

For certain applications, such as retrofitting an older room other materials are employed in shielding (or could be relevant to Monte Carlo shielding problems). The composition (by mass fraction) and density of some such materials are presented in Table 21.2.

21.3.2 Primary Radiation and TVLs

21.3.2.1 Direct Calculation

Directly calculating HVLs and TVLs with Monte Carlo has many attractive features. In addition to providing better accuracy than analytic approaches, HVL/TVLs can also be examined under broad-beam versus narrow-beam geometries. This means that the HVL/TVLs can be determined for a variety of field sizes, which may be of particular interest for machines where the field size is limited, for example, Cyberknife.

A Monte Carlo model of the beam line of the accelerator is sufficient for the task of calculating transmission of the primary beam in a barrier. Such a Monte Carlo model can be readily validated against attenuation (such as using percent-depth dose measurements in water). In place of developing a Monte Carlo model of the accelerator head (e.g., beginning with electrons incident on a bremsstrahlung target), it is also possible to

---

**Table 21.1** Elemental Composition (By Mass Fraction) of Various Types of Concrete

<table>
<thead>
<tr>
<th>Type</th>
<th>Density (g/cm³)</th>
<th>H</th>
<th>C</th>
<th>O</th>
<th>Na</th>
<th>Mg</th>
<th>Al</th>
<th>Si</th>
<th>S</th>
<th>K</th>
<th>Ca</th>
<th>Ti</th>
<th>V</th>
<th>Cr</th>
<th>Mn</th>
<th>Fe</th>
<th>Ba</th>
</tr>
</thead>
<tbody>
<tr>
<td>Portland</td>
<td>2.30</td>
<td>0.010</td>
<td>0.001</td>
<td>0.529</td>
<td>0.016</td>
<td>0.002</td>
<td>0.034</td>
<td>0.337</td>
<td>0.001</td>
<td>0.013</td>
<td>0.044</td>
<td>0.054</td>
<td>0.003</td>
<td>0.002</td>
<td>0.014</td>
<td>0.463</td>
<td></td>
</tr>
<tr>
<td>Ordinary</td>
<td>2.35</td>
<td>0.006</td>
<td>0.001</td>
<td>0.498</td>
<td>0.017</td>
<td>0.003</td>
<td>0.046</td>
<td>0.315</td>
<td>0.001</td>
<td>0.019</td>
<td>0.083</td>
<td>0.016</td>
<td>0.001</td>
<td>0.001</td>
<td>0.012</td>
<td>0.012</td>
<td></td>
</tr>
<tr>
<td>Magnetite</td>
<td>3.53</td>
<td>0.003</td>
<td></td>
<td>0.331</td>
<td></td>
<td>0.009</td>
<td>0.026</td>
<td>0.100</td>
<td></td>
<td>0.048</td>
<td>0.071</td>
<td></td>
<td>0.003</td>
<td></td>
<td>0.474</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Barytes</td>
<td>3.35</td>
<td>0.004</td>
<td></td>
<td>0.312</td>
<td></td>
<td>0.001</td>
<td>0.023</td>
<td>0.010</td>
<td></td>
<td>0.001</td>
<td>0.050</td>
<td></td>
<td>0.001</td>
<td></td>
<td>0.474</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Magnete and Steel</td>
<td>4.64</td>
<td>0.002</td>
<td></td>
<td>0.138</td>
<td></td>
<td>0.004</td>
<td>0.004</td>
<td>0.016</td>
<td></td>
<td>0.001</td>
<td>0.056</td>
<td></td>
<td>0.001</td>
<td></td>
<td>0.758</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Limonite and Steel</td>
<td>4.54</td>
<td>0.007</td>
<td></td>
<td>0.156</td>
<td></td>
<td>0.004</td>
<td>0.006</td>
<td>0.015</td>
<td></td>
<td>0.001</td>
<td>0.058</td>
<td></td>
<td>0.001</td>
<td></td>
<td>0.755</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Serpentine</td>
<td>2.1</td>
<td>0.016</td>
<td></td>
<td>0.512</td>
<td>0.001</td>
<td>0.135</td>
<td>0.019</td>
<td>0.209</td>
<td></td>
<td>0.004</td>
<td>0.068</td>
<td>0.16</td>
<td></td>
<td></td>
<td>0.031</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

begin simply with the final photon field (skipping the transport through the head). Historically, this approach was used to calculate TVL values based on monoenergetic photon beams. This is clearly only an approximate solution. Contemporary studies use polyenergetic spectra. For example, the Mohan spectra (1983) have been used by several groups to calculate the TVL of barriers (Avila-Rodriguez et al., 2005; Facure et al., 2007). This process allows the user to avoid developing even a beam line model of the accelerator, simplifying the modeling as well as the radiation transport. However, it is a less robust modeling approach as it does not account for the specifics of individual treatment heads or changes with treatment parameters. Of note is the change in TVL with field size (Jaradat and Biggs 2007), particularly for 60Co and low-energy x-ray beams. Such evaluation would be particularly useful when designing shielding for a vault with a dedicated stereotactic machine with a limited aperture size.

When the primary beam is obliquely incident on a barrier, the required barrier thickness is less than the calculated value (based on normal incidence) due to the obliquity of the beam. For angles of incidence less than 45°, the effective thickness of the barrier (i.e., the physical thickness times the obliquity factor) is given by \( t / \cos(\theta) \), where \( t \) is the normal barrier thickness and \( \theta \) is the angle of incidence of the primary beam. For angles of incidence greater than 45°, this adjusted obliquity factor thickness may not be adequate because more photons are likely to scatter normal to the barrier and, hence, may not travel the oblique path through the concrete. NCRP methodology specifies rules of the thumb to correct this effect by adding one or two HVLs to \( t \), depending on the beam energy and shielding material. This was found by the experimental work of Kirn et al. (1954) and verified by the Monte Carlo work of Biggs (1996).

### 21.3.2.2 Derivation of General Parameters

For primary beam shielding considerations, determination of the TVL\(_i\) (the first TVL) and TVL\(_e\) (the equilibrium TVL) are useful calculations because they allow for arbitrary barriers to be evaluated. In this case, there is little difference between the direct shielding calculation approach (Section 21.3.2.1) and the general parameter approach to determine the shielding metrics of TVL\(_i\) and TVL\(_e\). Of note, the divergence of the beam must be removed from the calculated TVLs so that they can be applied to a barrier at an arbitrary distance from the source. Perhaps, the earliest publication on primary beam spectra and TVLs was that of Nelson and LaRiviere (1984). Their calculation, based on EGS, was generic, that is, a simple target, and did not include specific beam line elements.

### 21.3.2.3 Inverse Square Law

McGinley (2001) made measurements of the exposure at various distances beyond a primary barrier for several linear accelerator vaults, expecting to see the dose fall off as the inverse square of the distance between the target and the point of interest. He was surprised to find that the effective source position was somewhat between the target and the barrier. He attributed this effect to photon scattering in the primary barrier. Biggs (2002) investigated this effect using the MCNP program and found that indeed the exposure fell off more rapidly than the inverse square, indicating that the effective photon source was not the target. The importance of this may be minimal because the experimentally derived TVLs as well as those using the Monte Carlo method have this factor inherently built into them, based, as they are, on single-point measurements. However, if multiple points beyond a barrier are being examined and compared, some discrepancy from an inverse square relationship would be expected.
21.3.3 Patient-Scattered Radiation Fractions and TVLs

Patient scatter is important only for the secondary barrier adjacent to the primary barrier and it should be noted that this radiation is incident obliquely on the secondary barrier. NCRP Report No. 151 recommends against using an obliquity factor for scattered radiation based on the diffuse nature of the scattering source.

21.3.3.1 Direct Calculation

With a validated primary photon beam, the radiation scattering from a patient/phantom and striking and penetrating a secondary barrier can be calculated using the Monte Carlo procedure (Biggs and Styczynski 2008). The scattering body should be appropriately considered as its size, shape, and composition can affect the amount of scattered radiation. Larger scattering bodies may scatter more radiation, but they will also attenuate more of the scattered radiation that is produced within them. Similarly, the size of the radiation field should be considered as larger field sizes will generate more scatter and the scatter will originate at different locations relative to the edge of the phantom (altering the amount of attenuation of the scattered radiation).

21.3.3.2 Derivation of General Parameters

Patient scatter fractions (NCRP Report No. 151 based on Taylor et al. (1999, 2000)) are the ratio of the scattered kerma to the dose at $d_{\text{max}}$ in a water sphere. For an arbitrary radiation beam, this approach can be applied to determine scattered radiation fractions and TVLs (e.g., Kry et al., 2009a). Such values can then be applied to general shielding problems. While this approach is well suited for solving shielding problems for unique radiation beams, it leaves substantial uncertainties for conditions such as irregularly shaped treatment vaults where a simple scatter fraction may not be sufficient information.

As with the direct calculation approach, care should be given to the size of the phantom and the radiation field (scatter is typically reported per unit area of the field, although a maximum field size would also be a reasonable approach).

Owing to an extended source size, the NCRP Report No. 151 specifically prohibits the use of an obliquity factor for scattered radiation. However, the use of an obliquity factor for scattered radiation is used in one of the examples. Because of this, Biggs and Styczynski (2008) used the MCNP code to determine the TVLs for 30° scattered radiation obliquely incident on the secondary barrier. Their findings showed that in contradistinction to NCRP recommendations, an obliquity factor could be justified.

21.3.4 Wall-Scattered Radiation Fractions and TVLs

Wall scatter is generally the least important component in determining secondary barrier or door thickness requirements.

21.3.4.1 Direct Calculation

As with patient scatter, once a primary beam is validated, radiation, as it scatters off the treatment vault walls, can be modeled relatively easily in a Monte Carlo calculation. Calculation of radiation scattered around a vault is likely to be largely limited by the geometrical definition of the vault and the number of particle histories.

It is unreasonable and unnecessary to model every detail of a clinical treatment vault, such as cupboards, sinks, and other ancillary equipment. However, large, dense structures should be modeled, or at least investigated to determine their importance in scattering radiation. Additionally, it is important to note the difference between the locations of the surface of the wall and the surface of the underlying concrete. Particularly in the ceiling, the concrete may be a substantial distance away from the false ceiling. Construction blueprints are generally a good source of information of the shielding-critical geometry.

Because wall-scattered radiation may have undergone multiple scatters, and usually has traveled large distances from the source, the particle flux is very low. Variance reduction will almost certainly be necessary to account for contributions to the shielding burden of radiation scattered throughout the vault. This may be a challenging step, and as with all variance reduction, care must be used to not suppress potentially relevant contributions to scatter throughout the treatment vault.

12.3.4.2 Derivation of General Parameters

General wall-scatter parameters can be readily calculated by impinging the primary beam on a wall (usually at both 90° (normal incidence) and 45°) and scoring the quantity and quality of the radiation scattered as a function of the angle.

Raso (1963) first investigated the scattering of gamma rays up to 10 MeV using the Monte Carlo approach for various angles of incidence and slab thicknesses. He demonstrated the sharp fall-off in scattered dose with increasing angle of scatter with respect to normal incidence. He also investigated inplane and out-of-plane scatter. Lo (1992) determined the scattering albedo for 4, 10, and 18 MV x-ray beams normally incident on ordinary concrete, iron, and lead using the ITS code (ITS 2.1). His test results from $^{60}$Co and $^{137}$Cs agreed very well with measured values. Two interesting results developed from this paper. The first is that the thickness for achieving saturation in the number of reflected photons is about 47 g/cm² indicating deep penetration before backscattering out of the medium. The second is that the average energy of the backscattered photons from concrete is around 0.15 MeV, whereas for iron it is ~0.25 MeV and for lead ~0.55 MeV, independent of incident energy over a wide range of megavoltage energies.

There has been little Monte Carlo work regarding second scatters, as would be important for maze calculations. Only direct maze calculations using Monte Carlo have been performed by Biggs (1991) and Al-Affan (2000). In the case of Biggs, the calculation was performed to validate the NCRP calculations for primary beams hitting the maze wall, whereas Al-Affan calculated...
for the more general maze situation. In both cases, they determined the x-ray fluence spectra striking the maze door (see Section 21.3.6).

21.3.5 Leakage Radiation and TVLs

Head leakage can be the most important component in defining secondary barrier thickness, particularly for treatment vaults that use IMRT. It is also particularly important for critical secondary barriers such as direct-shielded doors; however, it is also typically the most challenging component of the photon field to calculate due to the complex and irregular nature of the head shielding, and the corresponding spatial variations in leakage fluence and spectra.

21.3.5.1 Direct Calculation

Although there have been Monte Carlo models of accelerators designed with extensive detail of the entire treatment head (Kase et al., 1998, Howell et al., 2005, Kry et al., 2007, Bednarz and Xu, 2009), no model has yet been built that has been validated for simulation of the leakage radiation from a linear accelerator head. Given the importance of head leakage in defining the secondary barrier shielding requirements, careful and well-validated Monte Carlo modeling is necessary. For such a simulation, in-field validation of an accelerator model would clearly not be sufficient to verify accurate head leakage calculations. Most commonly, an analytic solution is sought for this part of the shielding question.

21.3.5.2 Derivation of General Parameters

Because it is hard to simulate head leakage levels accurately, general parameters for analytic models tend to be crudely defined. Because head leakage levels are mandated for accelerator manufacturers to be less than 0.1% of the primary beam, this is typically taken as the relative dose rate of head leakage (although it may be scaled by the amount of modulation used in IMRT treatments). Nelson and LaRiviere (1984) performed simple calculations to determine the penetration of leakage radiation in concrete, using a uniform spherical lead shell as a shielding model. Kase et al. (2003) and Li et al. (2006) performed TVL measurements for leakage radiation and found good agreement with Nelson and Lariviere’s (1984) calculations. Consequently, although it has not been employed to date, a simple Monte Carlo model (e.g., a simple lead shell) may therefore be sufficient for calculating the shielding burden of leakage radiation. Of course, the acceptability of this approach depends on the desired level of accuracy.

21.3.6 Mazes

There is relatively little Monte Carlo work performed on mazes in low-energy rooms where neutrons are not a problem, suggesting that analytical calculations suffice, at least for standard room geometries. Al-Affan (2000) used the MCNP package to calculate the dose and energy spectrum of photons at the door of a maze for a 6 MV bremsstrahlung beam. He showed that the photon energy spectrum at the door reached values up to 300 keV, but the most probable energy was <100 keV. However, although he presented the dose at the door in comparison with the dose at the isocenter (given in units of Gy/photon), he did not compare the results with the standard NCRP calculation. Biggs (1991) calculated the dose at the door of a maze for 4 and 10 MV x-ray beams (ignoring neutrons) using the ITS Monte Carlo code. Two types of rooms were considered: (i) a conventional room where the gantry rotation axis is parallel to the inner maze wall and (ii) a room where the primary beam is directly incident on the outer maze wall. It is difficult to calculate the dose in the latter configuration under the NCRP formalism because the coefficients are not well defined in this circumstance. He found that the Monte Carlo results were in agreement with the NCRP calculation at 10 MV for the conventional room, but a factor of 2 lower at 4 MV. For the nonstandard room, the Monte Carlo predicted a lower value by a factor of 3 for 4 MV and 1.5 for 10 MV. He also showed that the most probable photon energy at the door was 70–80 keV, irrespective of configuration, although the nonstandard room had a larger high-energy component.

21.3.7 Direct Door Shielding

There has been little to no interest in using the Monte Carlo technique to aid the calculation of direct door shielding for low-energy machines. There are two reasons for this. The first is that the most important source of radiation for direct-shielded doors is leakage radiation and a successful model that describes the leakage fluence and energy as a function of polar and azimuthal angles has not yet been developed. Second, because neutrons are not involved in the calculation, the “door edge” effect (Figure 2.10, NCRP Report No. 151) is much less problematic as only lead is used for shielding and door overlap can be generous.

21.3.8 HVAC Ducts

Depending on the size of the rooms and the required air flow rate, the heating, ventilation and air conditioning (HVAC) ducts can be quite large and are therefore a source of radiation leakage to a point just outside the barrier. Monte Carlo investigation of HVAC ducts is therefore of interest, particularly for rooms without mazes. However, no work on this subject has yet been conducted. In an unpublished work, Biggs and Styczynski found that for a typical HVAC duct located in the nonprimary plane, the 10th-value attenuation factor is approximately 60–70 cm, depending on the location and primary beam energy. Further study of this issue is warranted.

* This effect refers to the overlap region between the shielding barrier and the door. Because the door is very thick for high-energy direct-shielded doors, rays passing obliquely through the barrier/door junction will see a reduced shielding thickness, and therefore result in an elevated dose at the edge of the door. While this can be solved by making the door considerably wider, the weight and cost escalate prohibitively and would require a much heavier duty door controller.
21.4 Neutrons from High-Energy Photon Therapy

Neutrons are relevant for the shielding of accelerators operated at 10 MV or above. Because neutrons undergo multiple reflections from the walls, the vault must be included in the calculation. Calculations may be more sensitive to other ancillary equipment, which has the potential to degrade the neutron energy.

21.4.1 Neutron Production

Photon production of neutrons has a minimum threshold energy that varies with atomic number. For example, $^{208}$Pb has a threshold of 7.4 MeV, $^{56}$Fe has a threshold of 11.2 MeV, and $^{184}$W has a threshold of 7.4 MeV. Thus, given the standard photon energies offered by the vendors, beam energies must be at least 10 MV for photoneutrons to be produced in lead and 15 MV for steel. However, even at high energies, the main concern is the door of the vault or HVAC ducts in a mazeless room. Other barriers (primary and secondary barriers) that adequately shield photons will also generally shield the neutrons.

Neutrons are produced primarily through photon interactions with high-Z materials in the accelerator head (Howell et al., 2005). Once generated, these neutrons largely exit into the treatment vault and scatter throughout the room as high-Z materials are not efficient at reducing their energy or fluence. Neutrons scatter and will cross the treatment vault $>2$ times, on average, before finally being captured (NCRP Report No. 79). Neutrons lose energy primarily through elastic scatter interactions with low-Z materials, particularly hydrogen. Once thermalized, neutrons may be captured. Neutron capture cross sections vary greatly between elements, but are particularly high for boron and cadmium. After the capture of the thermal neutron, a photon is emitted by the excited nucleus. These gamma rays are typically of high energy; for example, after capturing a neutron, boron emits a photon of $E = 2.4$ MeV. Neutron shielding must therefore address all these stages of neutron absorption. First, thermalization of the neutrons (by, e.g., polyethylene), followed by absorption of the neutrons (by, e.g., a boron layer), and followed by attenuation of the emitted capture gamma rays (by, e.g., lead). This layered approach to neutron shielding is standard in the doors of radiotherapy treatment vaults.

Neutron source strength values vary between accelerator manufacturers and nominal energies. Followill et al. (2003) provide the broadest overview of source strengths, while Howell et al. (2009) provide detailed spectral measurements for common types and energies of accelerators currently available. Electron therapy has less neutron production than photon therapy because $(e,n)$ cross sections are lower than those for $(\gamma,n)$, and because electron therapy target currents are less than those for photon therapy (Lin et al., 2001). Electron-only accelerators, such as mobile linear accelerators used for intraoperative radiation therapy, are an exception to this, but there has been no Monte Carlo work to date on neutron production from these machines.

21.4.2 Neutron Spectra

Many different codes and modeling approaches have been used to study neutrons produced by medical accelerators. Agosteo et al. (1993) calculated the spatial distribution of the neutron flux inside an accelerator vault and confirmed the results using the MCNP code. However, their results did not lead to a practical application of neutron fluence calculation. Mao et al. (1997) and Kase et al. (1998) simulated a detailed head model of a clinical accelerator and calculated the production of neutrons and their transport through the head shielding into a typical therapy vault using EGS for photon production and MORSE for neutron production. They determined the neutron spectra outside the primary beam and identified the sources of neutrons in the accelerator head. Zanini et al. (2004) used MCNP4B-GN, which includes a routine to model photoneutron production, to investigate in-field neutron spectra for various collimation configurations. They only modeled the main components of the head, although they included a multileaf collimator (MLC). For an 18 MV linac, the most probable neutron energy was $\sim 500$ keV with a sharp drop in the spectrum above 1 MeV. Garnica-Garza (2005) also modeled only the main head components of a 15 MV linac using MCNP4C to generate the in-field spectra. The spectrum of room-degraded neutrons, with an average energy of 0.35 MeV was far different from the primary spectrum, with an average energy of 1.6 MeV. Howell et al. (2005) compared measured and calculated neutron spectra for an 18 MV linac, completely modeling the linac head. In this case, there was good agreement for a closed MLC field, but, for nonzero fields, the agreement was relatively poor. Barquero et al. (2005) modeled the accelerator head in MCNPX as a ball of tungsten with an isotropic neutron source. Facure et al. (2006) used a similar approach to study the scattered and thermal neutron fluence inside a room with the MCNP code. There are three publications from an Iranian group (Zabihzadeh et al., 2009, Ghiasi and Mesbahi 2010, Mesbahi et al., 2010). In the first paper, they used the same approach as Facure et al. (2006) and showed spectra of scattered neutrons for 10, 15, and 18 MV x-ray beams. Their spectra showed the typical fast and thermal neutron peaks. In the other two papers, the main components of the linac head were simulated. The first of these two papers examined the in-room spectra for 18 MV for three field sizes, while the second was primarily concerned about neutrons in the maze, but showed a neutron spectrum at 1 m from the source in the absence of room shielding.

Several authors have looked at the impact of field size on neutron flux or dose equivalent. Interestingly, seemingly contradictory results have been found. Mao et al. (1997) found that the neutron yield decreased with increasing field size, consistent with the decreased fluence calculated by Facure et al. (2006). Garnica-Garza (2005) found an increase in neutron yield with increasing field size, consistent with Kry et al. (2009b) who calculated an increase in dose equivalent with increasing field size. These disparate results likely reflect differences in neutron yield (neutrons produced in the head) versus flux in the treatment vault.
Caution must be used in interpreting some Monte Carlo results. Many authors have used simple models of accelerators for simulating neutrons, including only beam line structures. However, because of the extensive production and scattering processes throughout the accelerator head, studies have found that this approach produces results that may be in sizeable error as compared to results determined with complete models of the accelerator head (Kase et al., 1998).

In summary, there has been considerable effort to determine the neutron spectra from high-energy linacs over the past 20 years, but there is not the same degree of agreement with measurements as there is with primary photon beams. This is, in no small part, due to the difficulties in measuring neutrons, and also due to limitations in many of the simulations. It is also difficult to compare various results since each evaluates neutron spectra under different conditions. The prime need for having the neutron spectra is to assess direct door shielding, HVAC ducts for mazeless rooms, and neutron dose at the end of a maze. There is clearly room for further research.

### 21.4.3 Barriers

Concrete, by virtue of its high hydrogen content, is a particularly good neutron shielding material. In addition to the variations in aggregate distribution as described in Section 21.3.1, Monte Carlo simulations of concrete may pose an additional challenge for neutrons because, as concrete ages, the moisture level decreases as it dries out, which can change the hydrogen content. Concrete can lose 1–2% of its weight through lost water due to evaporation, affecting its ability to attenuate fast neutrons (Chilton 1984). Unfortunately, it is very difficult to account for this phenomenon and it is routinely neglected in Monte Carlo simulations, thereby adding to the overall uncertainty in calculations. Kase et al. (2003) has most thoroughly evaluated neutron TVLs for various concretes. They found reasonable agreement with measurements.

When space becomes an issue, either in new room construction or in room remodeling, high-atomic-number materials such as lead and steel are frequently used as shielding material to conserve space. However, for photon beams with energies >10 MV, neutrons will be produced where the primary beam strikes the high-Z barrier material. Some of these neutrons will subsequently produce capture gamma rays having an energy greater than 2 MeV. McGinley and Butker (1994) investigated this effect experimentally and derived an empirical formula for the photoproduction of neutrons from lead at 15 and 18 MV and for steel at 18 MV. They also showed that the contribution of capture gamma rays can be conservatively estimated to be 1.7 times the transmitted photon dose. However, these data were restricted to existing vaults. Thus, there is no data for steel and concrete laminated barriers at 15 MV and no data at all at 10 MV (though note that since the photoneutron threshold for steel is 11.2 MeV (see Section 21.4.1), there is no neutron production on steel for 10 MV beams). Facure et al. (2008) sought to remedy this by performing Monte Carlo calculations using MCNP5 for a bremsstrahlung beam with a peak energy of 11 MeV. They were able to show that steel generates no photoneutrons since the reaction thresholds are too high. They were also able to show that lead generates no substantial capture gamma rays, but there is sizeable neutron production in lead. Because McGinley’s 1.7 factor is very conservative (he took the worst case and there was considerable spread in the data), a Monte Carlo approach should be able to determine the dependence of this factor on other parameters such as thickness and relative location of the lead.

### 21.4.4 Mazes

The first empirical formula to determine the neutron dose at the end of a maze was due to Kersey (1979). This formula, or a variant of it, has been recommended by the NCRP (NCRP Report No. 79, 1984). Several authors have used the Monte Carlo method to solve this problem. The first difficulty is to generate a neutron spectrum that represents the neutron fluence coming from the head. Agosteo et al. (1995) simplified the concept by employing a point copper target at the center of a hollow tungsten sphere and adopting the method of Tosi et al. (1991) for the photoneutron spectrum. Carinou et al. (1999) used the same approach for generating the neutron spectrum and used this to investigate the neutron dose at various points along the maze and compared the results with empirical models. The results showed that the Monte Carlo results predicted lower values than the Kersey approach. Falcao et al. (2007) used the MCNP Monte Carlo code to examine the neutron fluence along the maze of multiple rooms housing accelerators with energies of 15, 18, and 25 MV. They also used Tosi’s technique to generate the neutron spectrum using a 10 cm tungsten sphere. However, unlike Carinou et al., they found that the Kersey method appeared to substantially underestimate the dose at the door. However, it was later shown (Wang et al., 2011) that this was due to a discrepancy in the ICRP conversion factors. Wu and McGinley (2003) modified the Kersey formula based on measurements along the maze of several typical linear accelerator rooms.

High-energy accelerators also have photon components to the maze dose. McGinley et al. (2000) evaluated the relative contributions of photon sources to the dose at the end of a maze, considering capture gamma rays, leakage, and scattered radiation. They found that capture gamma rays were generally the dominant source of photon dose, and that empirical equations could overestimate the leakage and scattered radiation by as much as a factor of 2.9.

### 21.4.5 Direct Door Shield (High Energy)

There has been little work done in this area, despite the pressing need for a Monte Carlo approach to this problem. The reason
for this need is that this is a difficult problem to solve using
hand calculations and usually requires considerable time and
effort to trace both photon and neutrons through the ends of
the door, the so-called door edge effect (see Figure 2.10, NCRP
Report No. 151). For Monte Carlo calculations, there is no ade-
quate model to describe the spectrum of leakage photons from
the head of the machine—unlike scattered radiation—as this
depends very heavily on the shielding design of the head and is
both vendor and model dependent. In addition, the spectrum
of neutrons has to be taken into account and, while the neu-
tron fluence spectrum has been derived, there is no assurance
that the same spectrum will apply at angles of 45° to 135° to
the beam axis. This is an area, which if solved by Monte Carlo,
would bear considerable fruit since it would not only speed up
calculations but it would also permit fast comparison between
different door designs.

### 21.4.6 HVAC Ducts

As with low-energy machines, this is an important problem for
mazeless rooms. For high-energy rooms, the problem is exac-
terbated by the need to shield against neutrons. Depending on
the location of the ducts, Monte Carlo input needs both photon
(leakage and scatter) and neutron spectra at the location of the
duct for different gantry angles. Even for rooms with mazes,
NCRP Report No. 151 provides no guidance in terms of ana-
lytical calculations, instead relying on the empirical evidence
of McGinley (NCRP Report No. 151). This empirical evidence
states that for an 18 MV linear accelerator, a dose equivalent
reduction of four for neutrons and two for photons will be pro-
duced by a 1.2-m-long duct wrap consisting of 2.5 cm Borated
polyethylene (BPE) and 1-cm lead in a 3.6-m-long maze. It is
clear that this data cannot be extrapolated easily to a general
situation and therefore a Monte Carlo-based approach would
allow some generalizations to be made. The merit of Monte
Carlo simulations for such situations is well known; the NCRP
formalism for handling neutrons from high-energy linear
accelerators was historically based on a large number of Monte
Carlo simulations performed by McCall.

### 21.5 Application of Monte Carlo for
Nonstandard Room Geometries

The NCRP-recommended calculation technique for door
shielding for rooms with mazes (NCRP Report Nos. 49, 51,
and 151) assumes a model where the primary beam does not
impinge upon the outer maze wall. In fact, the assumption is
made that the primary beam does not point in the direction
of the maze wall (see, e.g., Figure 2.8, NCRP Report No. 151).
However, for many rooms, this is not the case. While this may
be a simpler calculation because only one scatter calculation is
needed, the dose albedo for photons incident on a concrete sur-
face and scattering at near 90° is not well known and is sharply
dependent on the scattering angle (Chilton 1984). Thus, a large
uncertainty in the value for the wall scatter component (α) may
exist for these nonstandard rooms. Biggs applied the Monte
Carlo approach to this problem and compared it with NCRP
calculations (Section 21.3.5).

Facure et al. (2010) looked at unusual beam geometries in a
standard room, such as orienting the beam toward the maze
(this would be a possible orientation for a Cyberknife machine,
for example). They were able to show that, although the Monte
Carlo calculations agreed with NCRP calculations for standard
geometries (the latter being the more conservative), “where the
axis of gantry rotation is redirected at, for example, 45° with
respect to the walls of the room, the photon doses at the entrance
can reach values up to seven times higher than those obtained
under the standard conditions, depending on the energy of the
primary beam.”

Other situations where nonstandard geometries may arise
include cases where the gantry rotation axis is parallel to the
maze wall (standard geometry), but the maze opening is very
large (>2.0–2.5 m). In such a case, one must integrate the NCRP
equations over this broad geometry—something that may be
handled more easily using Monte Carlo.

Finally, there are rooms that have a curved maze entrance
or the vault itself has an unusual configuration or the gantry
rotation plane is not parallel to any wall. Again, NCRP calcula-
tions can be performed, but with difficulty, and a Monte Carlo
approach would have an advantage.

### 21.6 Nonstandard Beams and Accelerators

The C-arm linac is the most common device used to deliver
radiotherapy. Correspondingly, these machines have been most
thoroughly studied. Nevertheless, there are novel variants of the
x-ray beams from such devices, as well as several other modes of
delivering radiation therapy. These scenarios are less common
clinically but are often of greater interest because there is less
documentation on how to evaluate the shielding requirements.
The applications of Monte Carlo to solving these specific prob-
lems are detailed below.

FFF beams are available on many new models of C-arm
accelerators. These beams have substantially different spectra
and profiles (Vassiliev et al., 2006), and therefore have different
shielding characteristics. Kry et al. (2009a) used BEAMnrc to
reproduce analytic parameters for shielding C-arm FFF linacs
to mimic the data in the NCRP 151 report (including primary
beam TVL, patient scatter fractions, etc.). They also evaluated
neutron production in these beams using MCNPX and found an
~80% reduction in dose equivalent per course of 18 MV IMRT
with a Varian accelerator (Kry et al., 2008). They solved for the
shielding of a low-energy and mixed-energy accelerator in a typ-
ical radiotherapy vault and found reduced thickness of primary
and secondary barriers by 10–20% compared to traditional beams.
These simulations were confirmed by measurements by
Vassiliev et al. (2007).
Radiosurgery uses small treatment fields, which are of particular interest because they do not meet the broad-beam assumption of the NCRP-151 guidelines. This issue was explored by Rodgers (2007) for Cyberknife beams incident on concrete, lead, and steel, who found that the TVL for these beams changes substantially with field size, particularly for field diameters smaller than 8 cm where the TVL can be more than 20% less than the broad-beam TVL.

Brachytherapy has moved toward high-dose rate procedures in dedicated bunkers. Analytic models exist for calculating concrete wall thicknesses (ICRP Report No. 33, NCRP Report No. 49); however, these exist only for specific shielding materials and sources. Additionally, as with most analytic models, maze and vault door shielding are more uncertain. Lymperopoulou et al. (2006) and Papagiannis et al. (2008) used GEANT4 to expand the available data by calculating attenuation through four TVLs for a wide variety of both shielding materials and brachytherapy sources. Perez-Calatayud et al. (2004) used GEANT4 to evaluate the kerma and x-ray spectra in an high dose rate (HDR) maze and at the vault door. They also compared their results to analytic solutions using albedo coefficients, showing generally good agreement, but also some cases where the analytic solution underestimated the dose by 30%.

Neutron and hadron therapy facilities have also received attention for bunker shielding. Chen et al. (2008) used two coupled Monte Carlo codes to solve for dose and neutron and photon spectra in the room, maze, and outside the bunker of a boron neutron capture therapy (BNCT) facility. Porta et al. (2005) did similar analysis of shielding requirements for a hadron therapy bunker.

Proton therapy facilities are becoming increasingly popular. While NCRP Report No. 144 and Report No. 51 address shielding issues for such facilities, the novelty and variety of these units make additional evaluations essential. This is particularly true because the neutron fluence and spectral data are much more challenging to determine and therefore more scarce for proton therapy facilities than for high-energy photon beams. Nevertheless, this spectrum is very important in determining correct shielding values (Avery et al., 2008). Kim et al. (2005) therefore conducted a Monte Carlo shielding evaluation of their facility under construction to confirm their analytic solution. Of interest, while the analytic solution typically overestimated the dose, this was not always the case. Several other Monte Carlo studies have investigated the neutron spectra associated with proton therapy in detail, and treatment parameters that affect the spectra, particularly as they pertain to vault shielding. Zheng et al. (2007, 2009) showed that the neutron spectra consisted of two main peaks, one at ~1 MeV (isotropic neutrons produced by evaporation processes) and a high-energy peak at the maximum proton energy (forward-peaked neutrons produced by inanuclear cascade reactions), and related neutron doses to treatment parameters (2007). Perez-Andujar et al. (2009) showed that the bulk of these neutrons were produced in the range-modulator wheel in passive scattering systems. Agosteo (2009) simulated neutron spectra secondary to proton therapy, but also investigated the vault maze and room activation. As proton therapy continues to evolve, shielding considerations will require more refinement and additional study. For example, Fan et al. (2007) evaluated shielding requirements for a laser-accelerated proton therapy system based on a unique proton production process. Proton shielding will certainly be an area of active research for many years.

## 21.7 Summary and Future Directions

Numerous Monte Carlo codes have been successfully used to conduct shielding evaluations of radiotherapy treatment vaults. Such studies have answered questions about direct shielding needs and capabilities, they have tested analytic models, and they have been used to develop new analytic models. However, there is still a great need for further Monte Carlo research in this area. More refined models of leakage and neutron spectra are important for solving problems in mazeless rooms, such as direct door and HVAC duct shielding, and would eliminate long, tedious calculations and ray tracing by generating simplified rules. In addition, modeling the beams from newer, non-standard accelerators, such as Cyberknife, tomotherapy, or FFF linear accelerators would be of substantial benefit.

## References


Biggs P.J., Obliquity factors for $^{60}$Co and 4, 10 and 18 MV x rays for concrete, steel and lead and angles of incidence between 0° and 70°. *Health Phys.* 70:527–536;1996.


FIGURE 4.6

Clinac 2100C, Type II applicator, 15 MeV beam

Varian Clinac 2100C, 15 MeV electron beam

FIGURE 5.2

Varian Clinac 2100C linear accelerator

FIGURE 6.5
FIGURE 8.3

FIGURE 11.5
FIGURE 12.5

FIGURE 12.7
FIGURE 16.5

(a) MatriXX measurement
(b) TPS calculation
(c) EPID measurement
(d) Monte Carlo simulation

FIGURE 16.10
FIGURE 17.4

MC dose  MC PET  MC PET + washout  Measurement

FIGURE 17.7

FIGURE 17.8
Modern cancer treatment relies on Monte Carlo simulations to help radiotherapists and clinical physicists better understand and compute radiation dose from imaging devices as well as exploit four-dimensional imaging data. With Monte Carlo-based treatment planning tools now available from commercial vendors, a complete transition to Monte Carlo–based dose calculation methods in radiotherapy could likely take place in the next decade. Monte Carlo Techniques in Radiation Therapy explores the use of Monte Carlo methods for modeling various features of internal and external radiation sources, including light ion beams.

The book—the first of its kind—addresses applications of the Monte Carlo particle transport simulation technique in radiation therapy, mainly focusing on external beam radiotherapy and brachytherapy. It presents the mathematical and technical aspects of the methods in particle transport simulations. The book also discusses the modeling of medical linacs and other irradiation devices; issues specific to electron, photon, and proton ion beams and brachytherapy; and the optimization of treatment planning, radiation dosimetry, and quality assurance.

Useful to clinical physicists, graduate students, and researchers, this book provides a detailed, state-of-the-art guide to the fundamentals, application, and customization of Monte Carlo techniques in radiotherapy. Through real-world examples, it illustrates the use of Monte Carlo modeling and simulations in dose calculation, beam delivery, kilovoltage and megavoltage imaging, proton radiography, device design, and much more.