Biological and Medical Physics, Biomedical Engineering

Oleg N. Vassiliev

Monte Carlo Methods for Radiation Transport

Fundamentals and Advanced Topics

Springer

BIOLOGICAL AND MEDICAL PHYSICS, BIOMEDICAL ENGINEERING

BIOLOGICAL AND MEDICAL PHYSICS, BIOMEDICAL ENGINEERING

The fields of biological and medical physics and biomedical engineering are broad, multidisciplinary and dynamic. They lie at the crossroads of frontier research in physics, biology, chemistry, and medicine. The Biological and Medical Physics, Biomedical Engineering Series is intended to be comprehensive, covering a broad range of topics important to the study of the physical, chemical and biological sciences. Its goal is to provide scientists and engineers with textbooks, monographs, and reference works to address the growing need for information.

Books in the series emphasize established and emergent areas of science including molecular, membrane, and mathematical biophysics; photosynthetic energy harvesting and conversion; information processing; physical principles of genetics; sensory communications; automata networks, neural networks, and cellular automata. Equally important will be coverage of applied aspects of biological and medical physics and biomedical engineering such as molecular electronic components and devices, biosensors, medicine, imaging, physical principles of renewable energy production, advanced prostheses, and environmental control and engineering.

(Continued on next page)

Stuart M. Lindsay, Department of Physics and Astronomy, Arizona State University, Tempe, Arizona, USA

David Mauzerall, Rockefeller University, New York, New York, USA

Eugenie V. Mielczarek, Department of Physics and Astronomy, George Mason University, Fairfax, Virginia, USA

Markolf Niemz, Medical Faculty Mannheim, University of Heidelberg, Mannheim, Germany

V. Adrian Parsegian, Physical Science Laboratory, National Institutes of Health, Bethesda, Maryland, USA

Linda S. Powers, University of Arizona, Tucson, Arizona, USA

Earl W. Prohofsky, Department of Physics, Purdue University, West Lafayette, Indiana, USA

Tatiana K. Rostovtseva, NICHD, National Institutes Health, Bethesda, Maryland, USA

Andrew Rubin, Department of Biophysics, Moscow State University, Moscow, Russia

Michael Seibert, National Renewable Energy Laboratory, Golden, Colorado, USA

David Thomas, Department of Biochemistry, University of Minnesota Medical School, Minneapolis, Minnesota, USA

More information about this series at http://www.springer.com/series/3740

Oleg N. Vassiliev

Monte Carlo Methods for Radiation Transport

Fundamentals and Advanced Topics

Springer

Oleg N. Vassiliev
Department of Radiation Physics
The University of Texas
MD Anderson Cancer Center
Houston, TX, USA

ISSN 1618-7210 ISSN 2197-5647 (electronic)
Biological and Medical Physics, Biomedical Engineering
ISBN 978-3-319-44140-5 ISBN 978-3-319-44141-2 (eBook)
DOI 10.1007/978-3-319-44141-2

Library of Congress Control Number: 2016954643

Printed on acid-free paper

This Springer imprint is published by Springer Nature
The registered company is Springer International Publishing AG
The registered company address is: Gewerbestrasse 11, 6330 Cham, Switzerland

In memory of my parents

Then are you so certain that your roulette playing will get us out of our difficulties?

F. Dostoevsky, “The Gambler”

Preface

This book is intended as an introductory graduate-level text on the application of the Monte Carlo method to radiation transport problems. The target audience is radiation medical physicists: students, faculty members, and researchers specializing in radiotherapy physics, medical imaging, or nuclear medicine. The book should be of interest to clinicians as well, because Monte Carlo-based software, no longer confined to the research environment, is gradually finding its way into routine clinical practice.

The types of problems that are important in the field of medical physics determined the material that was selected for the book. Rather than focusing on the practical application of Monte Carlo techniques, however, the book focuses on the fundamentals of the method: its mathematical foundations, the numerical techniques on which it relies, its optimization strategies, and the statistical aspect of its calculations. With this approach, most of the information is quite general, and parts should be useful to a broad audience. More advanced topics are included as well, such as the adjoint formulation of the transport problem, the transport of charged particles in an external magnetic field, microdosimetry, elements of stochastic transport theory, and grid-based solvers. Inclusion of these topics makes the text more complete and extends the book into areas of recent significant developments.

An important objective of this book is to introduce the basic concepts, terminology, and formalism of radiation transport theory. This material, of course, is necessary to understand how transport problems are solved with the Monte Carlo method. It is also of significant interest in its own right because it is the basis for methods other than Monte Carlo, analytical and numerical, that have been used extensively in radiation medical physics. Several such methods are covered in the book.

Our didactic approach reflects the view expressed by N. Metropolis and S. Ulam in their seminal paper "The Monte Carlo method" (1949) that Monte Carlo is a "statistical approach to the study of differential equations, or more generally, of integro-differential equations." The equation that we study in this book is the

Boltzmann transport equation. For this reason, we dedicate an entire chapter to the equation and its various forms. Only after the equation is explained do we introduce algorithms for solving it.

The chapters and appendix of the book can be summarized as follows:

- Chapters 1 and 2 present a general introduction to the Monte Carlo method with an emphasis on sampling techniques, an essential element of any Monte Carlo algorithm. Sampling techniques are used to generate random numbers and vectors that have distributions required by the algorithm.
- Chapter 3 begins with definitions of the fundamental quantities of radiation transport theory, such as cross sections, free path, and fluence. Next is a rather elementary introduction to the Boltzmann equation followed by examples of its various forms. We conclude the chapter with more advanced topics: a general algorithm for solving the Boltzmann equation with the Monte Carlo method and the related topic of biasing techniques, which together form the mathematical basis for algorithm optimization.
- Chapter 4 discusses three main components of a Monte Carlo algorithm for radiation transport problems: generation of a particle trajectory, tallying, and variance reduction. Tallying is the process of deriving a numerical estimate of a quantity of interest from information contained in particle trajectories. Here, and throughout the book, the word estimate is used instead of calculate because Monte Carlo is a statistical method. This by no means implies poor accuracy of the result. Variance reduction is a broad term referring to a variety of optimization methods that reduce statistical uncertainties without introducing systematic error or bias.
- Chapter 5 is dedicated to the transport of charged particles such as electrons, protons, and heavy ions. Most Monte Carlo algorithms for charged particles rely on multiple scattering models. We cover all the classic models for energy loss fluctuations (energy straggling), angular distribution, and transverse and longitudinal spatial displacements. This chapter also includes sections on transport in magnetic fields and the charge exchange process, which is particularly important near the end of a heavy ion track.
- In the last two chapters, Chaps. 6 and 7, we present two advanced topics: microdosimetry with elements of stochastic transport theory and grid-based solvers of the Boltzmann equation. The calculation of microdosimetric characteristics is a problem fundamentally different from more conventional problems, such as the dose calculation, because the Boltzmann equation is not applicable in this case. For this reason, in this chapter, we introduce another equation, the stochastic transport equation, and discuss algorithms for solving it. Grid-based Boltzmann equation solvers are deterministic algorithms that present a viable alternative to Monte Carlo. The best-known algorithm of this type is Acuros (Vassiliev et al. 2010), which was translated into the clinic almost instantly, for treatment planning for radiotherapy of cancer. The grid-based Boltzmann equation solver, however, remains a relatively new technology and has the potential for improvement and for use in new applications. In Chap. 7, we explain step-by-step how an algorithm of this type works.

- Appendix A provides a summary of the concepts and methods of probability theory and statistics to help the reader better understand the material presented in the book and the statistical nature of the Monte Carlo method. In Appendix B, some of the mathematics used in the book is clarified.

Acknowledgments

The author thanks I.A. Matveeva (M.Arch.) and E.A. Butovskaya (M.Arch.) who created the first version of most of the illustrations in the book. Their work defined the overall style and served as a template for further revisions.

Houston, TX, USA Oleg N. Vassiliev

Reference

Vassiliev, O.N., Wareing, T.A., McGhee, J., Failla, G., Salehpour, M.R.: Validation of a new grid-based Boltzmann equation solver for dose calculation in radiotherapy with photon beams. Phys. Med. Biol. **55**(3), 581–598 (2010)

Contents

Chapter 1
Introduction

1.1 The Monte Carlo Method

The Monte Carlo method was the name suggested by Metropolis in the late 1940th for a statistical approach to solving neutron diffusion and multiplication problems that was being developed at that time at the Los Alamos Laboratory (Metropolis 1987). The approach was outlined in a letter sent by von Neumann to the leader of the Theoretical Division, Richtmyer (Richtmyer and von Neumann 1947). The letter stated that the principle of the statistical method was suggested by Ulam. However, this was not the first known application of the method now known as Monte Carlo to radiation transport. According to Sergè (2007) and Anderson (1986), Fermi "invented" the method and used it to study moderation of neutrons more than a decade earlier, in 1934, at the University of Rome. He, however, has not published anything on the technique he used (Sergè 2007).

What exactly is the Monte Carlo method? Metropolis and Ulam (1949) described the method as "a statistical approach to the study of differential equations, or more generally, of integro-differential equations that occur in various branches of the natural sciences." Since then, the application of the Monte Carlo method has been extended beyond differential and integro-differential equations, and beyond the natural sciences. Clearly, Metropolis and Ulam viewed Monte Carlo as a numerical method for solving mathematical problems, equations, including those that arise in physics. This means that before applying the Monte Carlo method to a physical problem, the problem should be formulated in mathematical terms, as an equation or a system of equations.

The uniqueness of the Monte Carlo method is in its statistical approach. It relies on statistical methods, such as sampling and inference. That does not mean that the method is applicable only to problems that involve random events or processes. Although, from the early days of Monte Carlo research, the method has been used predominantly to study random phenomena. Radiation transport being an inherently random process is a perfect example. In contrast to classical statistics, in the Monte

O.N. Vassiliev, *Monte Carlo Methods for Radiation Transport*, Biological and Medical Physics, Biomedical Engineering, DOI 10.1007/978-3-319-44141-2_1

Carlo method samples are generated by a computer algorithm rather than collected from observations. With respect to statistical inference, in Monte Carlo algorithms the solution of a mathematical problem is found as an estimate of a parameter of a distribution, usually the mean, or, more generally, of a given function of the parameter. Hence, in the Monte Carlo method, the numerical solution has a statistical uncertainty, in addition to any systematic errors that the algorithm may introduce. The statistical uncertainty is determined by the width of the distribution of the estimate of the parameter. Uncertainties are reported as a confidence interval, or simply as the standard deviation of the estimate. Statistical uncertainties normally decrease with increasing sample size.

The general Monte Carlo schema for solving a physical problem can be outlined as follows.

1. The problem is formulated in mathematical terms. This involves developing a model of the phenomenon, which leads to an equation or system of equations for the quantity of interest.
2. A statistical interpretation of the problem is formulated, where the quantity of interest is expressed as a parameter of a distribution, for example the mean, or a function thereof.
3. An algorithm for sampling the distribution is developed.
4. Estimators for the parameter and its statistical uncertainty are derived. These can be as simple as the sample average and the sample variance. In radiation transport calculations estimators are often called tallies.
5. The algorithm and the estimators are optimized to reduce the computing time needed to achieve the desired level of statistical uncertainties. Methods for achieving this goal without introducing systematic error, or a bias, are referred to broadly as variance reduction methods.
6. A sample is generated, that is sufficiently large to achieve the desired level of statistical uncertainty of the estimate of the parameter.
7. The parameter and its uncertainty are estimated using the sample.

Later in this chapter we provide a few simple examples that illustrate these steps. For random phenomena, the statistical interpretation of the problem may be immediately obvious. For example, to calculate the dose distribution from a radiation source, we would sample initial parameters of a particle from the distribution that characterizes the source, then generate a particle trajectory by copying on a computer the actual process, and as we do so, we would tally the dose this particle delivers. By generating a sufficiently large number of trajectories, we would achieve the desired level of statistical uncertainties in the calculated dose distribution. This approach is called analogous simulation. Its advantage is the simplicity, and clear physical interpretation of each step of the algorithm. The drawback is the lack of the flexibility necessary for optimizing the algorithm. Some optimization techniques can still be applied, but without a mathematical formulation of the problem, these techniques can only be based on intuitive arguments and therefore are unlikely to maximize the performance of the algorithm, and may even be erroneous.

1.2 The Monte Carlo Method in Radiation Medical Physics

The use of the Monte Carlo method in research and for solving applied problems of radiation medical physics has increased over the past two-three decades at a rate that can, perhaps, be described as unprecedented. The number of published papers on the subject has been increasing (Rogers 2006; Seco and Verhaegen 2013) at a rate much exceeding the rate of growth of all records in scientific literature databases, such as Medline (Larsen and von Ins 2010). Although the number of publications is a poor metric, because "it completely ignores the scientific impact of each paper" (Stallings et al. 2013), the vast number of papers does reflect the multitude and wide range of problems solved with the Monte Carlo method, and the persistently high level of activity in this area.

The interest in Monte Carlo techniques has grown from the need for accurate numerical methods for solving a variety of radiation transport problems that arise in radiation therapy, medical imaging, and nuclear medicine. These problems range from calculating parameters characterizing the performance of an ion chamber, to the calculation of patient dose distributions for radiotherapy treatment planning, or doses to patients and personnel from imaging devices, such as a computed tomography scanner. For a review of the practical applications of the Monte Carlo method in radiation therapy, we refer to the recently published book edited by Seco and Verhaegen (2013). We add that with the rapid progress in hadron therapy, interest has been increasing in understanding biological effects of radiation, where the Monte Carlo method has also proven to be an indispensable tool (Nikjoo et al. 2008).

Treatment planning for radiotherapy of cancer is a good example of the complexity of computational problems encountered in modern medical physics. It requires the calculation of three-dimensional (3D) dose distributions in a volume that includes the tumor and surrounding tissues at a spatial resolution of approximately 2–3 mm. Current radiotherapy treatment techniques are very sophisticated. Photon beam therapy, for example, uses multiple beams that target the tumor from different directions, selected so as to avoid, when possible, irradiating vital organs. The intensity of each beam has a complex pattern of spatial and temporal variations, optimized to produce the best possible dose distribution and achieve the treatment objectives. Beam intensity is modulated by beam modifying devices, following the instructions generated by treatment planning software. One such device is the multileaf collimator, which dynamically blocks parts of the beam during treatment. Before reaching the patient, the beam undergoes scattering in several beam modulating and beam collimating devices, resulting in a complex angular, spatial and energy distribution of photons and secondary particles, mostly electrons. Radiation transport in the heterogeneous anatomy of the patient also poses a challenge, because radiation beams penetrate different materials: soft tissue, lung, bone, air cavities, etc. Regardless of the complexity of the problem, the American Association of Physicists in Medicine (AAPM) (AAPM Report 85,

2004) recommends that the patient dose calculation error should not exceed 2 %. Requirements on the computing time are also rather stringent. In routine treatment planning, calculation of a patient dose distribution cannot take longer than a few minutes. The reason is that achieving an optimal 3D dose distribution requires multiple iterations, performed either interactively or by optimization software, with the dose distribution recalculated at each iteration. Another difficulty is that patients cannot be completely immobilized during treatment delivery. For example, due to the respiratory motion, lung tumors constantly move. It is not unusual to see the range of tumor motion exceeding 1 cm. To achieve a high accuracy of the dose distribution in this case, dose calculations have to be performed separately for several, usually ten, 3D snapshots of the anatomy of the patient acquired at different phases of the respiratory cycle. The calculations are then combined to calculate the total dose delivered to the patient. While this procedure is not yet standard in clinical practice, it is quite common in the research setting.

The required level of accuracy of the computed dose can certainly be achieved with the use of the Monte Carlo method. Furthermore, having a solid mathematical basis, the method is highly reliable: large errors are unlikely, if calculations are set up correctly. In addition, the method is relatively simple, at least conceptually, which simplifies the entire process, from learning the technique, to software testing, to interpretation of numerical results. Finally, the progress made in the solving problems faced in the field of radiation medical physics using the Monte Carlo technique would not have been possible, at least not on the present scale, without the development of user-friendly software, which has made access to this powerful method much easier. The following software systems have been used extensively in radiation medical physics:

- EGS, "Electron-Gamma Shower" (Ford and Nelson 1978; Nelson et al. 1985; Kawrakow et al. 2013).
- BEAM (Rogers et al. 1995), a graphic user interface for EGS, designed specifically for modeling medical accelerators.
- GEANT, "Generation of Events ANd Tracks" (Brun et al. 1978; Agostinelli et al. 2003; Allison et al. 2006).
- Geant4-DNA (Incerti et al. 2010; Bernal et al. 2015), an extension of the general purpose toolkit Geant4 to very low energies, for modeling radiation damage on a microscopic scale.
- MCNP, "Monte Carlo N-Particle" (Hendricks and Briesmeister 1991; Brown 2003).
- PENELOPE, "PENetration and Energy LOss of Positrons and Electrons" (Baro et al. 1995; Salvat et al. 2011; Salvat 2015).

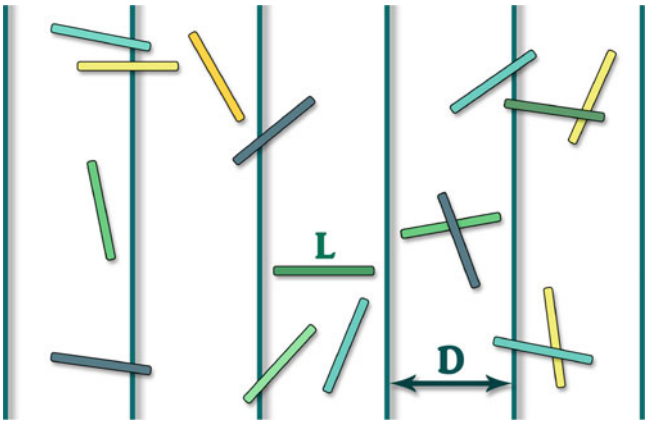

Fig. 1.1 Estimation of π

1.3 Estimation of π

It is very common for books on the Monte Carlo method to discuss the statistical experimental method for estimating π, first described by de Buffon (1777). This book is not an exception. Our purpose, however, is only to illustrate with this example the general schema of the Monte Carlo method presented above.

The earliest description of the method in English that we are aware of was given by Hall (1873). The idea of the method is as follows. Parallel lines separated by a distance D are drawn on a planar surface. We assume that the number and the length of lines are infinite. A straight line segment of length L is randomly placed on the planar surface (Fig. 1.1). The probability p that this segment intersects a parallel line is (Hall 1873)

$$p = \frac{2L}{\pi D}. \tag{1.1}$$

The probability p can be estimated experimentally. In the experiment described by Hall (1873), a fine steel wire was thrown on a wooden planar surface on which were drawn equidistant parallel lines. If the wire was thrown N(total) times, and it intersected a parallel line N(intersect) times, then the estimate of the probability of intersection in one trial is:

$$p\,(\text{experimental}) = \frac{N\,(\text{intersect})}{N\,(\text{total})}. \tag{1.2}$$

If we replace the exact probability p in Eq. (1.1) with its approximate value found experimentally, Eq. (1.2), and then solve the equation for π, we find an estimator for π, $\hat{\Theta}_\pi$:

$$\pi \approx \hat{\Theta}_\pi = \frac{2L}{p\,(\text{experimental})\,D}. \tag{1.3}$$

This simple example has all the elements of the general Monte Carlo schema. The statistical interpretation of the problem is that the quantity of interest, π, is expressed as a function of p. The observed N(intersect) has a binomial distribution (see Appendix A). Hence, π is expressed as a function of parameter p of the binomial distribution. The distribution is sampled experimentally by throwing a thin steel wire on a planar surface on which parallel lines are drawn and counting the number of times the wire intersects a line. We have also derived an estimator of π. It is given by Eq. (1.3). This estimator, however, has a serious flaw. If after N(total) trials, the wire has never intersected any of the lines, then $N(\text{intersect}) = 0$, and we have division by zero in Eq. (1.3). However small the probability of this occurrence may be, it is not exactly zero, it is a finite number. It is not a small number, if the lines are well separated, $D \gg L$. This shows that estimators are an important part of a Monte Carlo algorithm and their possible deficiencies should not be overlooked. In our example, the problem is that the estimate of π is biased, which means that the expectation of $\hat{\Theta}_\pi$ is not equal to π. In fact, the expectation is infinite. The variance of the estimate is also infinite, which makes analysis of uncertainties nontrivial.

Finally, the optimization parameter of this problem is the ratio L/D. In general, the optimization parameter should be chosen to minimize the variance of the estimate for a given number of trials. In our case the variance does not exist, therefore another measure of the width of the distribution of the estimate has to be minimized. Alternatively, the estimator can be modified so that the expectation and variance of the estimate both are finite. For this particular problem, however, using a non-Monte Carlo algorithm is a much better strategy. A historical review by Glaisher (1873), in the same issue of "Messenger of Mathematics" where the paper by Hall was published, gives a good insight into the methods for calculating π.

1.4 Calculation of Volumes

Problem

Calculate volume V, Fig. 1.2.

Statistical Interpretation of the Problem

If a random point has a uniform distribution in a volume V_B (bounding volume) such that $V \subseteq V_B$, (the entire volume V is inside V_B), then the probability p of this point falling within V is equal to the ratio of the volumes, $p = V/V_B$. If N is the total number of random points uniformly distributed in V_B, and N_V is the number of points that fall within V, then the distribution of N_V is the binomial distribution with parameter p. Hence, again, the quantity of interest is expressed as function of parameter p of the binomial distribution, $V = pV_B$.

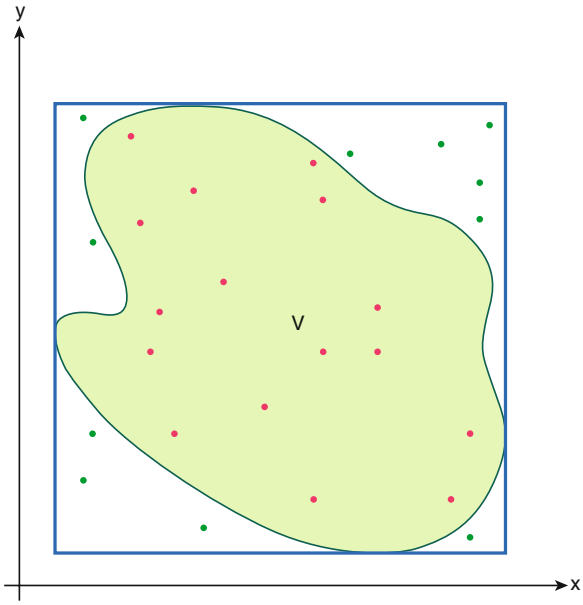

Fig. 1.2 Calculation of volumes. The bounding volume V_B is the rectangle

Sampling of the Distribution

To sample a random number N_V, we generate N random points with a uniform distribution in V_B and count the number of points N_V that fall within V. To simplify the sampling of a uniform distribution, a rectangular V_B is used in two dimensions, and an orthotope ("box") is used for higher dimensionalities. Sampling techniques are discussed in detail in the next chapter.

Estimator

For a sufficiently large N, we have $p \approx N_V/N$. Thus, the estimator of V is $\hat{\Theta}_V = V_B N_V/N$. This time the estimate is unbiased, i.e., $\mathrm{E}\{\hat{\Theta}_V\} = V$, where $\mathrm{E}\{.\}$ denotes the expectation.

Optimization

The optimization parameter for this problem is V_B. We, however, prefer to work with a dimensionless parameter, $x = V/V_B$, $0 \leq x \leq 1$. We need to find the variance of

the estimate as a function of x, and then find a value of x that minimizes the variance. Here we can use the formula for the variance of a binomial distribution

$$\mathrm{Var}\left\{\frac{V_B}{N}N_V\right\} = \frac{V_B^2}{N^2}\mathrm{Var}\left\{N_V\right\} = \frac{V^2}{N}\left(\frac{1}{x}-1\right). \tag{1.4}$$

Thus, to minimize the variance we need to maximize x, by choosing the smallest V_B. At this point we can also find the fractional uncertainty

$$\frac{1}{V}\sqrt{\frac{V^2}{N}\left(\frac{1}{x}-1\right)} = \frac{1}{\sqrt{N}}\sqrt{\frac{1}{x}-1}. \tag{1.5}$$

The inverse square root of the sample size N is very common for fractional uncertainties in Monte Carlo calculations.

Algorithm

1. Generate N random points uniformly distributed within the bounding volume, $N \gg 1$.
2. Count the number of points N_V that fall within volume V, Fig. 1.2.
3. Calculate the volume: $V \approx V_B N_V/N$.
4. Calculate the uncertainty using Eq. (1.4), Eq. (1.5), or the sample variance [Appendix A, Eq. (A.58)].

1.5 Calculation of Integrals

Problem

Calculate the integral

$$I = \int_a^b f\,(x)\,\mathrm{d}x\,;\;\; a > b. \tag{1.6}$$

Statistical Interpretation

Let $g(x)$ be an arbitrary function that satisfies conditions A and B

$$\text{A.}\quad \int_a^b g\,(x)\,\mathrm{d}x = 1\,. \tag{1.7}$$

$$\text{B.} \quad g(x) \geq 0; \quad \text{for } \forall\, x \in [a,b] \ . \tag{1.8}$$

We also set $g(x) = 0$ for x outside the integration interval, $[a, b]$. Then $g(x)$ can be interpreted as a probability density function of a random variable ξ, and the integral can be interpreted as the expectation value

$$\int_a^b f(x)\,\mathrm{d}x = \int_a^b g(x)\frac{f(x)}{g(x)}\,\mathrm{d}x = \mathrm{E}\left\{\frac{f(\xi)}{g(\xi)}\right\} . \tag{1.9}$$

Again, ξ is a random number with distribution $g(x)$.

Sampling the Distribution

We need to sample ξ from distribution $g(x)$. Given that $g(x)$ can be any function that satisfies conditions A and B, it should be chosen so that the sampling algorithm is simple and fast. The easiest to sample distribution is a uniform distribution in interval $[a, b]$. In that case $g(x)$ is a constant. However, we have not yet discussed optimization of the algorithm, which, as shown below, introduces another requirement that must be considered when choosing $g(x)$.

Estimator

According to Eq. (1.9), we need to estimate an expectation value. We can use for that purpose the sample average

$$I \approx \hat{\Theta}_I = \frac{1}{N}\sum_{i=1}^{N}\frac{f(\xi_i)}{g(\xi_i)}, \tag{1.10}$$

provided that $g(x)$ is chosen so that the ratio $f(x)/g(x)$ is finite for all $x \in [a, b]$.

Optimization

Here we present the importance sampling optimization technique. We need to find a function $g(x)$ that satisfies conditions A and B, and minimizes the variance of the ratio $f(\xi)/g(\xi)$. The variance is

$$\mathrm{Var}\left\{\frac{f(\xi)}{g(\xi)}\right\} = \int_a^b g(x)\left[\frac{f(x)}{g(x)}\right]^2 \mathrm{d}x - \left[\int_a^b g(x)\frac{f(x)}{g(x)}\mathrm{d}x\right]^2 . \tag{1.11}$$

In the second integral, $g(x)$ is cancelled out, which means that we need to minimize the first integral

$$\int_a^b g\,(x)\left[\frac{f\,(x)}{g\,(x)}\right]^2 \mathrm{d}x\,. \tag{1.12}$$

To find the minimum of this integral, we write an expression for the variance of the ratio $|f(\xi)|/g(\xi)$

$$\mathrm{Var}\left\{\frac{|f\,(\xi)\,|}{g\,(\xi)}\right\} = \int_a^b g\,(x)\left[\frac{f\,(x)}{g\,(x)}\right]^2 \mathrm{d}x - \left[\int_a^b |f\,(x)\,|\mathrm{d}x\right]^2. \tag{1.13}$$

Given that the variance must be nonnegative, Eq. (1.13) shows that the minimal possible value of the integral in Eq. (1.12) is

$$\left[\int_a^b |f\,(x)\,|\mathrm{d}x\right]^2. \tag{1.14}$$

This minimum is achieved, if we choose

$$g\,(x) = \frac{|f\,(x)\,|}{\int_a^b |f\,(x)\,|\mathrm{d}x}\,. \tag{1.15}$$

To prove that the minimum is reached, insert Eq. (1.15) into Eq. (1.12). It can also be shown that with $g(x)$ given by Eq. (1.15), the variance in Eq. (1.11) remains nonnegative. Thus, the variance of the ratio $f(\xi)/g(\xi)$ and therefore statistical uncertainty of the calculated integral are minimal, if we chose $g(x)$ given by Eq. (1.15). If the integrand is positive, i.e., $f(x) = |f(x)|$ for all $x \in [a, b]$, this choice produces a result with a zero variance. However, to implement this zero-variance algorithm one needs to know the integral I, because it is present in the denominator in Eq. (1.15). Thus, a more practical approach is to choose, if possible, a simple $g(x)$ similar to the expression on the right-hand side of Eq. (1.15).

Algorithm Groundwork

Implementation of Monte Carlo algorithms often requires some preparatory work or analysis. We refer to this step as "groundwork."

In case of the algorithm for calculating an integral, at this step we choose a function $g\,(x)$. The function must be normalized and nonnegative (conditions A and B). It also should be such that sampling from distribution $g(x)$ is simple and fast, and the ratio $f(x)/g(x)$ is finite. For maximum algorithm efficiency $g(x)$ should be similar to the right-hand side of Eq. (1.15). If the integrand function

$f(x)$ has singularities, it is recommended that they are included in $g(x)$, so that the ratio $f(x)/g(x)$ does not have singularities. Once a function $g(x)$ is selected, the algorithm can be finalized.

Algorithm

1. Generate a sample $\{\xi_1, \xi_2, \ldots, \xi_N\}$, where the distribution of ξ_i is $g(x)$; $N \gg 1$
2. Calculate the sample average of the ratio $f(\xi)/g(\xi)$:

$$\frac{1}{N}\sum_{i=1}^{N}\frac{f(\xi_i)}{g(\xi_i)}. \tag{1.16}$$

3. Output the result

$$I = \int_a^b f(x)\,\mathrm{d}x \approx \frac{1}{N}\sum_{i=1}^{N}\frac{f(\xi_i)}{g(\xi_i)}. \tag{1.17}$$

If the uncertainty of the result is needed, then the standard deviation of the sample average is also calculated.

Example 1

Calculate the integral

$$I = \int_0^\infty f(x)\,\mathrm{d}x = \int_0^\infty \sqrt{x}\exp(-x)\mathrm{d}x. \tag{1.18}$$

We choose

$$g(x) = \exp(-x), \quad x \geq 0, \tag{1.19}$$

and ensure that conditions A and B are satisfied:

$$\text{A.} \quad \int_0^\infty \exp(-x)\mathrm{d}x = 1. \tag{1.20}$$

$$\text{B.} \quad \exp(-x) > 0. \tag{1.21}$$

We also check that

$$\frac{f(x)}{g(x)} = \sqrt{x} \text{ is finite for } \forall\, x \in [0, \infty). \tag{1.22}$$

Algorithm

1. Generate a sample $\{\xi_1, \xi_2, \ldots, \xi_N\}$, where distribution of ξ_i is $\exp(-x)$; $N \gg 1$.
2. Calculate the sample average of the ratio $f(\xi)/g(\xi)$:

$$\frac{1}{N}\sum_{i=1}^{N}\sqrt{\xi_i}. \tag{1.23}$$

3. Output the result

$$I \approx \frac{1}{N}\sum_{i=1}^{N}\sqrt{\xi_i}. \tag{1.24}$$

Example 2

Calculate the integral

$$I = \int_0^1 f(x)\,\mathrm{d}x = \int_0^1 \frac{1}{\sqrt{x}}\exp(-x)\mathrm{d}x. \tag{1.25}$$

The integrand has a singularity, $1/\sqrt{x}$, at $x \to 0$. We choose a function $g(x)$ with the same asymptotic behavior for small x

$$g(x) = \frac{1}{2\sqrt{x}};\ \ 0 < x \leq 1. \tag{1.26}$$

Then we must verify that conditions A and B are satisfied

$$\text{A.} \quad \int_0^1 \frac{1}{2\sqrt{x}}\,\mathrm{d}x = 1. \tag{1.27}$$

$$\text{B.} \quad \frac{1}{2\sqrt{x}} \geq 0;\ \ \text{for } \forall\, x \in (0, 1]. \tag{1.28}$$

We also note that

$$\frac{f(x)}{g(x)} = 2 \cdot \exp(-x) \quad \text{is finite for } \forall\, x \in [0, 1]\,. \tag{1.29}$$

Algorithm

1. Generate a sample $\{\xi_1, \xi_2, \ldots, \xi_N\}$, where distribution of ξ_i is $2/\sqrt{x}$, $0 < x \leq 1$; $N \gg 1$.
2. Calculate the sample average of the ratio $f(\xi)/g(\xi)$:

$$\frac{1}{N} \sum_{i=1}^{N} 2 \cdot \exp(-\xi_i). \tag{1.30}$$

3. Output the result

$$I \approx \frac{1}{N} \sum_{i=1}^{N} 2 \cdot \exp(-\xi_i). \tag{1.31}$$

In all the algorithms that we have discussed so far, there is a step where we generate random numbers or random points that have a certain required distribution. In the next chapter we discuss a variety of methods for performing this task.

References

American Association of Physicists in Medicine (AAPM): Tissue inhomogeneity corrections for megavoltage photon beams. Radiation Therapy Committee Task Group 65. Report No. 85. Medical Physics Publishing, Madison, WI (2004)

Agostinelli, S., Allison, J., Amako, K., Apostolakis, J., Araujo, H., Arce, P., Asai, M., Axen, D., et al.: Geant4 – a simulation toolkit. Nucl. Inst. Methods A **506**(3), 250–303 (2003)

Allison, J., Amako, K., Apostolakis, J., Araujo, H., Dubois, P.A., Asai, M., Barrand, G., Capra, R., et al.: Geant4 developments and applications. IEEE Trans. Nucl. Sci. **53**(1), 270–278 (2006)

Anderson, H.L.: Scientific uses of the MANIAC. J. Stat. Phys. **43**(5/6), 731–748 (1986)

Baro, J., Sempau, J., Fernandez-Varea, J.M., Salvat, F.: PENELOPE: an algorithm for Monte Carlo simulation of the penetration and energy loss of electrons and positrons in matter. Nucl. Inst. Methods B **100**(1), 31–46 (1995)

Bernal, M.A., Bordage, M.C., Brown, J.M.C., Davidkova, M., Delage, E., El Bitar, Z., Enger, S.A., Francis, Z., et al.: Track structure modeling in liquid water: a review of the Geant4-DNA very low energy extension of the Geant4 Monte Carlo simulation toolkit. Phys. Med. **31**(8), 861–874 (2015)

Brown, F.B. (ed.): MCNP – a general Monte Carlo N-particle transport code, version 5. Los Alamos National Laboratory Report LA-UR-03-1987, Los Alamos, NM (2003)

Brun, R., Hagelberg, R., Hansroul, M., Lassalle, J.C.: Simulation program for particle physics experiments, GEANT. User Guide and Reference Manual. DD/78/2. CERN, Data Handling Division (1978)

de Buffon, G.-L.L.: Essai d'Arithm'etique Morale. Suppl'ement a' l'Histoire Naturelle **4**, 46–148 (1777)

Ford, R.L., Nelson, W.R.: The EGS code system: computer programs for the Monte Carlo simulation of electromagnetic cascade showers (Version 3). Stanford Linear Accelerator Center, Report 210, Stanford, CA (1978)

Glaisher, J.W.L.: Remarks on the calculation of π. Messenger Math. **2**, 119–128 (1873)

Hall, A.: On an experimental determination of π. Messenger Math. **2**, 113–114 (1873)

Hendricks, J.S., Briesmeister, J.F.: Recent MCNP developments. Los Alamos National Laboratory Report LA-UR-91-3456, Los Alamos, NM (1991)

Incerti, S., Ivanchenko, A., Karamitros, M., Mantero, A., Moretto, P., Tran, H.N., Mascialino, B., Champion, C., et al.: Comparison of Geant4 very low energy cross section models with experimental data in water. Med. Phys. **37**(9), 4692–4708 (2010)

Kawrakow, I., Mainegra-Hing, E., Rogers, D.W.O., Tessier, F., Walters, B.R.B.: The EGSnrc Code System: Monte Carlo simulation of electron and photon transport. Report PIRS-701: National Research Council of Canada (2013)

Larsen, P.O., von Ins, M.: The rate of growth in scientific publication and the decline in coverage provided by Science Citation Index. Scientometrics **84**, 575–603 (2010)

Metropolis, N.: The beginning of the Monte Carlo method. Los Alamos Sci. **15**(Special issue), 125–130 (1987)

Metropolis, N., Ulam, S.: The Monte Carlo method. J. Am. Stat. Assoc. **44**(247), 335–341 (1949)

Nelson, W.R., Hirayama, H., Rogers, D.W.O.: The EGS4 code system Report SLAC-265. Stanford Linear Accelerator Center, Report 265, Stanford, CA (1985)

Nikjoo, H., Uehara, S., Emfietzoglou, D., Brahme, A.: Heavy charged particles in radiation biology and biophysics. New J. Phys. **10**, 075006 (2008)

Richtmyer, R.D., von Neumann, J.: Statistical methods in neutron diffusion. Los Alamos Scientific Laboratory Report LAMS-551. Los Alamos (1947)

Rogers, D.W.O.: Fifty years of Monte Carlo simulations for medical physics. Phys. Med. Biol. **51**(13), R287–R301 (2006)

Rogers, D.W.O., Faddegon, B.A., Ding, G.X., Ma, C.M.: BEAM - A Monte Carlo code to simulate radiotherapy treatment units. Med. Phys. **22**(5), 503–524 (1995)

Salvat, F.: The PENELOPE code system. Specific features and recent improvements. Ann. Nucl. Energy **82**, 98–109 (2015)

Salvat, F., Fernandez-Varea, J.M., Sempau, J.: Penelope-2011: A Code System for Monte Carlo Simulation of Electron and Photon Transport. Issy-les-Moulineaux, OECD/NEA Data Bank, France (2011)

Seco, J., Verhaegen, F. (eds.): Monte Carlo Techniques in Radiation Therapy. CRC, Boca Raton, FL (2013)

Sergè, E.: From X-Rays to Quarks: Modern Physicists and their Discoveries. Dover, Mineola, NY (2007)

Stallings, J., Vance, E., Yang, J., Vannier, M.W., Liang, J., Pang, L., Dai, L., Ye, I., Wang, G.: Determining scientific impact using a collaboration index. Proc. Natl. Acad. Sci. USA **110**(24), 9680–9685 (2013)

Chapter 2
Sampling Techniques

2.1 Sampling a Uniform Distribution

Let us start with a uniform one-dimensional distribution. For a random number uniformly distributed within (0, 1) we use a special notation, γ. In all algorithms we will have a step: generate γ. The main tool for generating γ is a pseudorandom number generator that is a computer program based on a deterministic algorithm. The multiplicative congruential algorithm (Lehmer 1951) is the simplest algorithm for generating uniformly distributed pseudorandom numbers.

Algorithm

1. $\xi_{i+1} = a\xi_i \bmod m$.
2. $\gamma_{i+1} = \frac{\xi_{i+1}}{m}$.

The algorithm generates a sequence of pseudorandom numbers, $\gamma_1, \gamma_2, \ldots, \gamma_N$. The algorithm has three parameters that must be initialized before the algorithm can be executed: a, m, and ξ_0. All three parameters are positive integers; a is called the multiplier, m is the modulus, and ξ_0 is the seed. The parameters are chosen such that the sequence $\gamma_1, \ldots, \gamma_N$ satisfies statistical tests for randomness. Another requirement affecting the choice of the parameters is that the generator must have a very long period. Because the range of an integer variable is finite, eventually the generator returns to one of its previous states, and starts repeating exactly the same sequence of pseudorandom numbers. In other words, it has a finite period. Fortunately, modern algorithms achieve such long periods that reaching the end of a sequence before a calculation is completed would be unusual. Random number generators have been covered extensively in the literature. For further reading on this topic, we refer to Fishman (1996).

O.N. Vassiliev, *Monte Carlo Methods for Radiation Transport*, Biological and Medical Physics, Biomedical Engineering, DOI 10.1007/978-3-319-44141-2_2

Most software packages provide an option for generating a seed ξ_0 randomly, using for example a computer's clock. Alternatively, ξ_0 can be set by the user. The latter method allows the user to perform calculations with exactly the same random sequence, which is very convenient for software testing. Assuming that we have tools for generating random numbers uniformly distributed within the interval $(0, 1)$, we will now discuss how these numbers can be used to generate other distributions.

Problem

Generate a random number ξ distributed uniformly in an interval (a, b), $b > a$.

Algorithm

1. Generate γ.
2. $\xi = (b - a)\gamma + a$.

Problem

Place a uniformly distributed point within a three-dimensional rectangular box. The box spans from x_a to x_b in the x-direction, from y_a to y_b in the y-direction, and from z_a to z_b in the z-direction.

Algorithm

1. Generate γ.
2. $\xi = (x_b - x_a)\gamma + x_a$.
3. Generate γ.
4. $\eta = (y_b - y_a)\gamma + y_a$.
5. Generate γ.
6. $\zeta = (z_b - z_a)\gamma + z_a$.

Here, (ξ, η, ζ) are the (x, y, z) coordinates of the point. Note that the algorithm uses three different and statistically independent gammas, generated at steps 1, 3, and 5.

2.2 The Inversion Method for Sampling Continuous Distributions

Problem

Generate a random number ξ that has a continuous distribution $f(x)$.

Algorithm Groundwork

Find the cumulative distribution function (CDF):

$$F(x) = \int_{-\infty}^{x} f(x)\,\mathrm{d}x. \tag{2.1}$$

Algorithm

1. Generate γ.
2. Solve equation $\gamma = F(\xi)$ for ξ.

In other words,

$$\xi = F^{-1}(\gamma). \tag{2.2}$$

Figure 2.1 illustrates the algorithm.

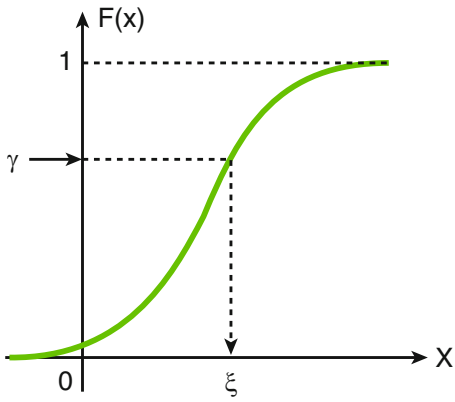

Fig. 2.1 Inversion method for a continuous distribution

Proof

Let us show that the algorithm, indeed, produces a random number ξ with distribution $F(x)$. By definition, $F(x) = P\{\xi \le x\} =$ [using Eq. (2.2)] $= P\{F^{-1}(\gamma) \le x\} =$ (because $F(x)$ is monotonic) $= P\{F\left[F^{-1}(\gamma)\right] \le F(x)\} =$ (using the definition of an inverse function)$= P\{\gamma \le F(x)\} =$ (because γ is uniformly distributed between 0 and 1, and $0 \le F(x) \le 1$ for any x) $= F(x)$. This proves that the algorithm samples ξ correctly. □

Example 1. Sampling a Chord-Length Distribution

This is an example from microdosimetry (see Chap. 6). A parallel beam of heavy charged particles, protons, for example, is incident on a sphere of unit diameter. Each particle entering the sphere travels a random path ξ (chord) within it. Assuming that the particles travel along a straight line, and do not stop within the sphere, the chord-length distribution is given by this simple formula

$$f(x) = \begin{cases} 2x, & \text{if } 0 \le x \le 1; \\ 0, & \text{otherwise.} \end{cases} \tag{2.3}$$

Write an algorithm for sampling the chord length ξ.

Algorithm Groundwork

First, we find the CDF:

1. For $x < 0$:

$$F(x) = \int_{-\infty}^{x} 0 \, dx' = 0. \tag{2.4}$$

2. For $x > 1$:

$$F(x) = \int_{-\infty}^{0} 0 \, dx' + \int_{0}^{1} 2x' \, dx' + \int_{1}^{x} 0 \, dx' = 1. \tag{2.5}$$

3. For $0 \le x \le 1$:

$$F(x) = \int_{0}^{x} 2x' \, dx' = x^2. \tag{2.6}$$

Then, we find the inverse of the CDF as in Eq. (2.2). The result is $\xi = \sqrt{\gamma}$.

Algorithm

1. Generate γ.
2. $\xi = \sqrt{\gamma}$.

Example 2: Sampling a Free Path

Neutral particles, such as photons, travel along a straight line between points of interaction with matter. The distance between interactions is called a free path. It is distributed exponentially (see Sect. 3.1):

$$f(x) = \begin{cases} \sigma \exp(-\sigma x), & \text{if } x \geq 0; \\ 0, & \text{if } x < 0, \end{cases} \tag{2.7}$$

where σ is the macroscopic total cross section (cm^{-1}) and x is the distance (cm). Write an algorithm for sampling the free path ξ.

Algorithm Groundwork

Again, first we need to find the CDF. For $x < 0$, we have $F(x) = 0$, and for $x \geq 0$:

$$F(x) = \int_0^x \sigma \exp\left(-\sigma x'\right) \mathrm{d}x' = 1 - \exp(-\sigma x). \tag{2.8}$$

Then, we solve Eq. (2.2) and find that in this case:

$$\xi = -\frac{1}{\sigma} \ln(1 - \gamma). \tag{2.9}$$

Because $1 - \gamma$ and γ have exactly the same distribution, we can slightly simplify the above formula:

$$\xi = -\frac{1}{\sigma} \ln \gamma. \tag{2.10}$$

Algorithm

1. Generate γ.
2. $\xi = -(1/\sigma) \ln \gamma$.

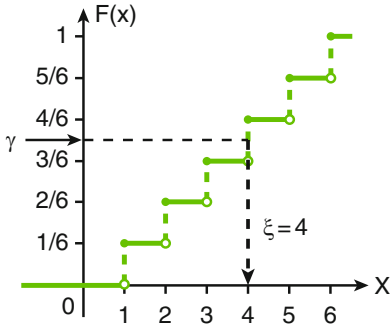

Fig. 2.2 Inversion of the CDF for one roll of a die

2.3 The Inversion Method for Sampling Discrete Distributions

The CDF inversion method is also applicable to discrete random variables. However, in this case the probability density $f(x)$ does not exist. Regardless, the CDF can always be found. For discrete random numbers the CDF is a stepwise-constant function. For example, the CDF for one roll of a die is shown in Fig. 2.2. This figure also shows how CDF inversion is performed in this case. Then, the algorithm for sampling a discrete random number ξ is:

Algorithm (See Fig. 2.2)

1. Generate γ.
2. Find where the dashed γ-line intersects the green line, $F(x)$.
3. From the intersect point, draw a vertical line and determine where it intersects the abscissa.
4. ξ is the point of intersection with the abscissa.

Figure 2.3 shows the flow chart for a generic algorithm based on the CDF inversion method for an arbitrary discrete distribution with the sample space $\{x_1, x_2, \ldots\}$, and respective probabilities $P_1, P_2, \ldots$.

For some important distributions a recurrence relation exists for the probabilities: $P_{k+1} = r(k) P_k$, where $r(k)$ is a known, usually simple, function. This simplifies the calculation of P_k in the second box of the flow chart.

Examples

Binomial and Poisson distributions. For a binomial distribution:

$$r(k) = \frac{n-k}{k+1} \frac{p}{1-p}, \tag{2.11}$$

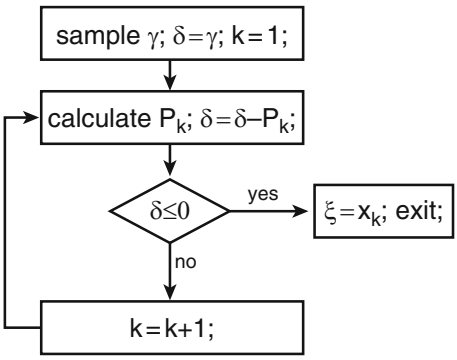

Fig. 2.3 A generic algorithm for sampling discrete distributions

and for a Poisson distribution:

$$r(k) = \frac{\mu}{k+1}. \tag{2.12}$$

Here n, p, and μ are parameters of the distributions.

2.4 Simple Rejection Method

The CDF inversion method is, in principle, applicable to any distribution, continuous or discrete. However, it is not necessarily the fastest or simplest. The simple rejection method is a good alternative that offers a very straightforward algorithm that, depending on the properties of the distribution, may or may not be more efficient than the CDF inversion method.

Algorithm Groundwork

Find the minimum bounding box for the probability density function (the red box in Fig. 2.4). If the box is infinite in any direction then, strictly speaking, this method does not work. In that case, the method can be used only approximately. If, for example, a tail of the distribution extends to infinity, the tail can be truncated. However, in this situation, we would recommend applying another sampling technique, either the inversion method or Neumann's method that is discussed in the next section.

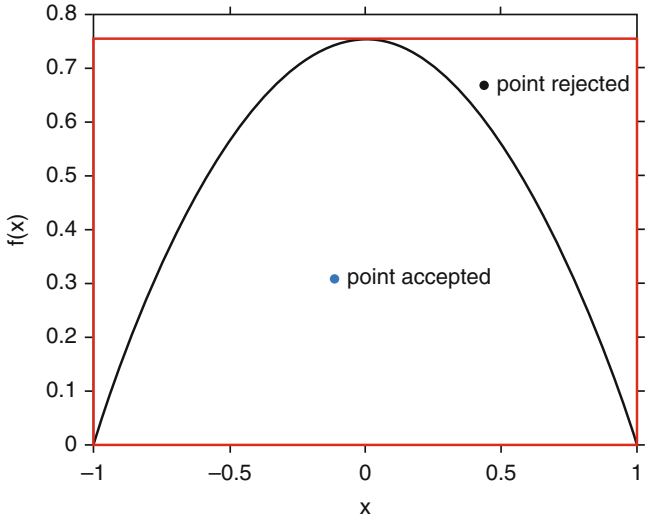

Fig. 2.4 Simple rejection method

Algorithm

1. Place a random point uniformly within the bounding box.
2. If the (x, y) coordinates of the point are (ξ, η), and $\eta < f(\xi)$, that is, the point falls below the $f(x)$ curve (see Fig. 2.4), then the point is accepted (go to step 3); otherwise, go back to step 1.
3. Output ξ.

The efficiency of the algorithm, defined as the probability of accepting a point, is equal to the ratio of the area under the $f(x)$ curve to the area of the bounding box, that is, 1/{area of the box}.

Proof

By definition,

$$F(x) = P\{\xi \leq x\}. \tag{2.13}$$

For this algorithm, the probability in the above equation is conditional, the condition being that the point was accepted. Then,

$$P\{\xi \leq x | \text{ accepted}\} = \frac{P\{\xi \leq x \text{ and accepted}\}}{P\{\text{accepted}\}}. \tag{2.14}$$

Using a total probability formula:

$$P\{\xi \le x|\text{ accepted}\} = \int_{x_{\min}}^{x_{\max}} \mathrm{d}x' \int_0^{y_{\max}} \mathrm{d}y' \times \frac{p\,(x', y')\, P\{\xi \le x \text{ and accepted}|x', y'\}}{P\{\text{accepted}\}}, \tag{2.15}$$

where the integral is over the bounding box, $p(x', y')$ is the probability density of the uniform distribution of the point location within the box. It is equal to 1/{area of the box} and therefore cancels out with $P\{\text{accepted}\}$. Finally, $P\{\xi \le x \text{ and accepted}|x', y'\} = 1$, if $x' \le x$ and $y' \le f(x')$; it is zero otherwise.
Then, the integral in Eq. (2.15) becomes

$$P\{\xi \le x|\text{ accepted}\} = \int_{x_{\min}}^{x} \mathrm{d}x' \int_0^{f(x')} \mathrm{d}y' = \int_{x_{\min}}^{x} \mathrm{d}x' f\left(x'\right) = F\left(x\right). \tag{2.16}$$

□

Example

Write an algorithm for sampling a random number ξ from the distribution:

$$f\left(x\right) = \frac{3}{4}\left(1 - x^2\right); \qquad -1 \le x \le 1\,. \tag{2.17}$$

The bounding box is: $-1 \le x \le 1$ and $0 \le y \le 3/4$; the efficiency is 1/{box area} = 2/3.

Algorithm

1. Sample ξ uniformly within $[-1, 1]$.
2. Sample η uniformly within $[0, 3/4]$.
3. If $\eta > (3/4)(1 - \xi^2)$, then the point is rejected, go back to step 1. Else, the point is accepted; output ξ.

2.5 Neumann's Method

Neumann's method is also a rejection algorithm. However, it offers much more flexibility than simple rejection.

Algorithm Groundwork

Choose a probability distribution $g(x)$ that is:

(a) similar to $f(x)$.
(b) simple, such that it is easy to sample from distribution $g(x)$; and
(c) a number $C \geq 1$ exists such that $Cg(x) \geq f(x)$ for any x. Choose $g(x)$ with the smallest C.

Algorithm

1. Sample ξ from distribution $g(x)$.
2. Generate γ.
3. If $\gamma \leq f(\xi)/[Cg(\xi)]$, then output ξ and exit. Else, go back to step 1.

If we choose $g(x) = \text{Const}$, then this algorithm is no different from simple rejection. If, on the other hand, we choose $g(x) = f(x)$, then $C = 1$ and steps 2 and 3 are not needed. The efficiency of the algorithm is defined as the probability of accepting a ξ, that is, $P\{\gamma \leq f(\xi)/[Cg(\xi)]\}$. First, let us show that the efficiency is equal to $1/C$. We start with a total probability equation:

$$P\{\gamma \leq f(\xi)/[Cg(\xi)]\} = \int_{-\infty}^{\infty} g(x)\,\mathrm{d}x \int_0^1 \mathrm{d}t\, P\{\gamma \leq f(\xi)/[Cg(\xi)]|x,t\}. \tag{2.18}$$

We notice that $P\{\gamma \leq f(\xi)/[Cg(\xi)]|x,t\} = 1$, if $t \leq f(x)/[Cg(x)]$, and 0 otherwise. Then, the integral can be calculated easily:

$$P\{\gamma \leq f(\xi)/[Cg(\xi)]\} = \int_{-\infty}^{\infty} g(x)\,\mathrm{d}x \int_0^{f(x)/[Cg(x)]} \mathrm{d}t = \frac{1}{C}. \tag{2.19}$$

□

Next, we prove that the algorithm samples ξ from distribution $F(x)$.

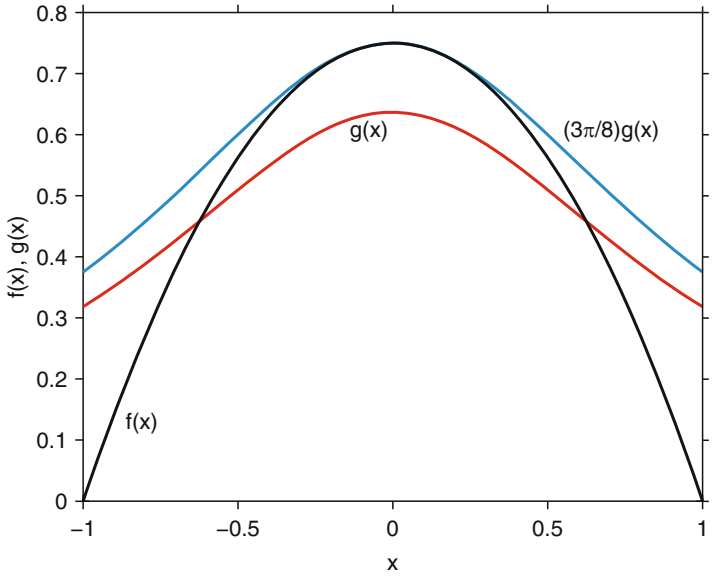

Fig. 2.5 Neumann's method

Proof

The proof is very similar to the proof of the simple rejection algorithm. The only difference is that at step 1, ξ is sampled from distribution $g(x)$ instead of from a uniform distribution. Then, obviously

$$F(x) = P\{\xi \le x | \text{ accepted}\} = C \int_{-\infty}^{x} g\left(x'\right) \mathrm{d}x' \int_{0}^{f(x')/[Cg(x')]} \mathrm{d}t$$
$$= \int_{-\infty}^{x} f\left(x'\right) \mathrm{d}x' = F(x)\,, \tag{2.20}$$

where we used $P(\text{accepted}) = 1/C$. □

Example

Write an algorithm for sampling a random number ξ from a distribution (Fig. 2.5):

$$f(x) = \frac{3}{4}\left(1 - x^2\right); \qquad -1 \le x \le 1\,. \tag{2.21}$$

.

Algorithm Groundwork

Choose

$$g(x) = \frac{2}{\pi\,(1+x^2)}. \tag{2.22}$$

In this case $C = 3\pi/8$ and the efficiency is approximately 0.85. Distribution $g(x)$ can be sampled by the CDF inversion method. The sampling formula is

$$\xi = \tan\left[\frac{\pi}{2}\left(\gamma - \frac{1}{2}\right)\right]. \tag{2.23}$$

Algorithm (See Fig. 2.5)

1. Sample ξ from distribution $g(x)$ using Eq. (2.23).
2. Generate γ.
3. If $\gamma \leq 1 - \xi^4$, then output ξ and exit. Else, go back to step 1.

2.6 Transformation of Random Variables

It may be possible to transform a random variable ξ into another variable $\eta = g(\xi)$ that has a distribution that is easier to sample. Then, to sample ξ one would, first, sample η and then perform the inverse transformation $\xi = g^{-1}(\eta)$.

If $f_\xi(x)$ is the distribution of ξ, then the distribution of η can be found as follows:

$$f_\eta(y) = f_\xi(x)\left|\frac{\mathrm{d}x}{\mathrm{d}y}\right| = \frac{f_\xi(x)}{|\mathrm{d}g(x)/\mathrm{d}x|}. \tag{2.24}$$

The right-hand side should be expressed as a function of y using $x = g^{-1}(y)$.

This method is also applicable to multidimensional variables. If ξ is an n-dimensional variable, and the transformation to another random variable is given by a set of functions g_i: $\eta_i = g_i(\xi_1, \ldots, \xi_n)$, $i = 1, \ldots, n$, then the distribution of the transformed variable is

$$f_{\eta_1,\ldots,\eta_n}(y_1, \ldots, y_n) = f_{\xi_1,\ldots,\xi_n}(x_1, \ldots, x_n)\left|\det\left\{\frac{\partial x_i}{\partial y_j}\right\}\right|, \quad i,j = 1, 2, \ldots, n, \tag{2.25}$$

where the last term on the right-hand side of the equation is the Jacobian.

Example

Sample distribution

$$f_\xi(x) = 2x\,\mathrm{e}^{-x^2},\ \ x \geq 0. \tag{2.26}$$

Instead, we will sample $\eta = \xi^2$. The distribution of η according to Eq. (2.24) is

$$f_\eta(y) = f_\xi(x)\,\frac{1}{2x} = \mathrm{e}^{-y}. \tag{2.27}$$

Algorithm

1. Sample η from distribution e^{-y}, $y \geq 0$. Use the CDF inversion method.
2. Output $\xi = \sqrt{\eta}$.

2.7 Sampling a Sum of Distributions

Problem

Sample distribution $f(x) = p_1 f_1(x) + p_2 f_2(x)$, where

$$p_1 \geq 0;\ \ p_2 \geq 0;\ \ f_1(x) \geq 0;\ \ f_2(x) \geq 0; \tag{2.28}$$

$$p_1 + p_2 = 1; \tag{2.29}$$

$$\int_{-\infty}^{+\infty} f_1(x)\,\mathrm{d}x = \int_{-\infty}^{+\infty} f_2(x)\,\mathrm{d}x = 1. \tag{2.30}$$

Algorithm

1. Generate γ.
2. If $\gamma < p_1$, then sample ξ from distribution $f_1(x)$. Else, sample ξ from distribution $f_2(x)$.

This algorithm can be generalized to a sum with any number of terms, including infinite sums:

$$f(x) = \sum_{i=1}^{\infty} p_i f_i(x)\,. \tag{2.31}$$

Algorithm

1. Sample a channel number i from the discrete distribution given by probabilities $\{p_1, p_2, \ldots\}$.
2. Sample ξ from distribution $f_i(x)$.

The algorithm follows from the observation that the right-hand side of Eq. (2.31) is the total probability equation, where $f_i(x)$ is the conditional distribution given that event i has occurred.

Example

Sample states of all particles after the interaction of a photon with an atom.

We have three channels:

1. Photoelectric absorption.
   - the photon is absorbed
   - a photoelectron is ejected
   - the atom is ionized
2. Compton scattering.
   - the photon momentum changes
   - a Compton electron is ejected
   - the atom is ionized
3. Pair production
   - the photon disappears
   - an electron and a positron are produced.

If x is the photon state before interaction, and $x_1, x_2, \ldots$ are states of all particles after the interaction, then we need to sample a distribution given by:

$$f(x \to x_1, x_2, \ldots) = p_1(x) f_1(x \to x_1, x_2, \ldots) + p_2(x) f_2(x \to x_1, x_2, \ldots) + p_3(x) f_3(x \to x_1, x_2, \ldots). \tag{2.32}$$

The algorithm is obvious. First, we sample the channel number: 1, 2, or 3. Or, in other words, we sample the type of interaction. Once the type of interaction is determined, we sample the parameters of particles after the interaction using the distribution function for the selected interaction type.

2.8 Compton Scattering

Example

Sample the scattering angle and energy of a photon after Compton scattering. We consider a simple model where the photon interacts with a free electron at rest. In this case, the differential scattering cross section is given by the Klein–Nishina formula (Butcher and Messel 1960; Attix 2004). In this model, the probability distribution of photon energy after scattering can be written as

$$f\left(E'\right) = A\left[\frac{E'}{E} + \frac{E}{E'} - \sin^2\theta\right], \tag{2.33}$$

where E and E' are the photon energies before and after the interaction, respectively, θ is the photon scattering angle, and A is the normalization constant. The initial photon energy, E, is given. The energy E' after the interaction is random, and we need to sample it. The scattering angle θ can be found as a function of E and E' from

$$E' = \frac{E}{1 + E\left(1 - \cos\theta\right)}, \tag{2.34}$$

where the energies are in the units of electron rest energy, $E = \text{energy}/m_e c^2$. The minimum and maximum E' can be found from Eq. (2.34) by setting $\theta = \pi$ and $\theta = 0$, respectively:

$$E'_{\min} = \frac{E}{1 + 2E}; \quad E'_{\max} = E. \tag{2.35}$$

We chose this example because Compton scattering is an important process and the sampling algorithm is nontrivial. It uses three techniques we introduced above: the CDF inversion, Neumann's method, and sampling a sum of two distributions. The earliest formulation of the algorithm that we were able to find was given by Butcher and Messel (1960). The algorithm is also described in the documentation for an early version of the EGS software system (Ford and Nelson 1978).

Algorithm Groundwork

First, we rewrite Eq. (2.33) in a slightly different form:

$$f\left(E'\right) = A\left(\frac{E'}{E} + \frac{E}{E'}\right)\left(1 - \frac{\sin^2\theta}{E'/E + E/E'}\right), \tag{2.36}$$

and then

$$f\left(E'\right) = A\left(\alpha_1 + \alpha_2\right)\left[\frac{\alpha_1}{\alpha_1 + \alpha_2} f_1\left(E'\right) + \frac{\alpha_2}{\alpha_1 + \alpha_2} f_2\left(E'\right)\right]\left(1 - \frac{\sin^2\theta}{E'/E + E/E'}\right), \tag{2.37}$$

where

$$f_1\left(E'\right) = A_1 E'; \quad f_2\left(E'\right) = \frac{A_2}{E'}, \tag{2.38}$$

and the normalization constants A_1 and A_2 can be found from:

$$A_1 \int_{E'_{\min}}^{E'_{\max}} E' \mathrm{d}E' = A_2 \int_{E'_{\min}}^{E'_{\max}} \frac{\mathrm{d}E'}{E'} = 1. \tag{2.39}$$

By comparing Eqs. (2.36) and (2.37), it can be determined that

$$\alpha_1 = \frac{1}{A_1 E}; \quad \alpha_2 = \frac{E}{A_2}. \tag{2.40}$$

Equation (2.37) is a basis for the implementation of the Neumann's method. To this end, we choose

$$g\left(E'\right) = \frac{\alpha_1}{\alpha_1 + \alpha_2} f_1\left(E'\right) + \frac{\alpha_2}{\alpha_1 + \alpha_2} f_2\left(E'\right), \tag{2.41}$$

and

$$C = A\left(\alpha_1 + \alpha_2\right). \tag{2.42}$$

It can be seen from Eq. (2.37) that with these choices the condition

$$Cg\left(E'\right) \geq f\left(E'\right) \tag{2.43}$$

is satisfied.

Algorithm Overview

1. Sample ξ from distribution $g\left(E'\right)$. Because $g\left(E'\right)$ is a sum of two distributions, it is sampled using the technique described in Sect. 2.7. Distributions, f_1 and f_2 are sampled using the inversion of the CDF method.
2. Generate γ.

3. If $\gamma < f(\xi)/[Cg(\xi)]$, then go to step 4. Else, go to step 1. Here $f(\xi)/[Cg(\xi)] = 1 - \sin^2\theta/(\xi/E + E/\xi)$, where $\sin\theta$ is calculated from Eq. (2.34).
4. Output $E' = \xi$, and the scattering angle θ that was calculated at step 3.

Algorithm

1. Generate γ.
2. If $\gamma < \alpha_1/(\alpha_1 + \alpha_2)$ then sample ξ from distribution f_1:

$$\xi = \sqrt{\left(E'_{\min}\right)^2 + 2\gamma/A_1}. \tag{2.44}$$

   Else, sample ξ from distribution f_2:

$$\xi = E'_{\min} \exp\left(\frac{\gamma}{A_2}\right). \tag{2.45}$$

   Of course, before Eq. (2.44) or Eq. (2.45) is used, a new γ is generated.
3. Calculate, using Eq. (2.34)

$$\sin^2\theta = 1 - \left(1 - \frac{1}{\xi} + \frac{1}{E}\right)^2. \tag{2.46}$$

4. Generate γ.
5. If $\gamma < 1 - \sin^2\theta/(\xi/E + E/\xi)$, then output $E' = \xi$ and θ, and exit. Else, go back to step 1.

2.9 Superposition Method

Problem

Sample a probability distribution defined as an integral:

$$f(x) = \int_{-\infty}^{+\infty} h(x,t)\,\mathrm{d}t. \tag{2.47}$$

Algorithm Groundwork

Choose a normalized and simple-to-sample distribution $g(t)$, and rewrite the integral as

$$\int_{-\infty}^{+\infty} h(x,t)\,\mathrm{d}t = \int_{-\infty}^{+\infty} g(t)\,\frac{h(x,t)}{g(t)}\,\mathrm{d}t. \tag{2.48}$$

The integral on the right-hand side is a form of the total probability equation. Then the ratio h/g can be interpreted as a conditional probability distribution, that is, the distribution of x given t:

$$\frac{h(x,t)}{g(t)} = f(x|\,t)\,. \tag{2.49}$$

Hence, the algorithm is:

Algorithm

1. Sample τ from distribution $g(t)$.
2. Sample ξ from distribution $f(x|\,\tau)$.

Example

Write an algorithm based on the superposition technique for sampling distribution:

$$f(x) = A\int_0^{\infty} t\,\exp\left[-t\,(x+t)\right]\mathrm{d}t; \quad x \geq 0; \quad A > 0. \tag{2.50}$$

Algorithm Groundwork

From the normalization condition

$$\int_0^{\infty} f(x)\,\mathrm{d}x = 1, \tag{2.51}$$

we find that

$$A = \frac{2}{\sqrt{\pi}}, \tag{2.52}$$

and write the integrand as a product of two distributions:

$$At \exp\left[-t\left(x+t\right)\right] = \left[t \exp\left(-tx\right)\right] \cdot \left[\frac{2}{\sqrt{\pi}} \exp\left(-t^2\right)\right]. \tag{2.53}$$

It is important to verify that both of these distributions are normalized:

$$\int_0^\infty t \exp\left(-tx\right) \mathrm{d}x = \frac{2}{\sqrt{\pi}} \int_0^\infty \exp\left(-t^2\right) \mathrm{d}t = 1. \tag{2.54}$$

Algorithm

1. Sample τ from the distribution $(2/\sqrt{\pi}) \exp\left(-t^2\right)$, $t \geq 0$. This is simply the positive half of a normal distribution. Sampling of the normal distribution is discussed in the next section.
2. Sample ξ from distribution $\tau \exp\left(-\tau x\right)$, $x \geq 0$. This the exponential distribution. It can be sampled by the inversion method.
3. Output ξ.

2.10 Sampling a Normal Distribution

We will use notation $N\left(\mu, \sigma^2\right)$ for a normal distribution with parameters μ, and σ^2:

$$f\left(x\right) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[-\frac{\left(x-\mu\right)^2}{2\sigma^2}\right]. \tag{2.55}$$

The general approach to sampling $N(\mu, \sigma^2)$ is, first, to sample the standard distribution, $N\left(0, 1\right)$ and then perform a simple transformation of the random number as shown below.

Algorithm

1. Sample δ from distribution $N\left(0, 1\right)$.
2. Output $\xi = \sigma\delta + \mu$.

Proof

The following shows that the above algorithm indeed produces a random variable with distribution $N(\mu, \sigma^2)$. We will use Eq. (2.24) for the distribution of a transformed random variable. In the following, $f_\delta(x)$ denotes the distribution of δ, which is $N(0,1)$, and $f_\xi(y)$ is the distribution of the transformed variable ξ. The transformation is $\xi = \sigma\delta + \mu$.

$$f_\xi(y) = f_\delta(x)\left|\frac{\mathrm{d}x}{\mathrm{d}y}\right| = f_\delta(x)\frac{1}{\sigma} = f_\delta\left(\frac{y-\mu}{\sigma}\right)\frac{1}{\sigma} = \frac{1}{\sqrt{2\pi\sigma^2}}\exp\left[-\frac{(y-\mu)^2}{2\sigma^2}\right]. \tag{2.56}$$

□

We can additionally show that the first two moments of ξ are correct:

$$\begin{aligned}\mathrm{E}(\xi) &= \mathrm{E}(\sigma\delta + \mu) = \sigma\mathrm{E}(\delta) + \mathrm{E}(\mu) = \sigma\cdot 0 + \mu = \mu.\\ \mathrm{Var}(\xi) &= \mathrm{Var}(\sigma\delta + \mu) = \sigma^2\mathrm{Var}(\delta) + \mathrm{Var}(\mu) = \sigma^2\cdot 1 + 0 = \sigma^2.\end{aligned} \tag{2.57}$$

Next, we need to address the problem of sampling the standard normal distribution, $N(0,1)$.

Method 1

An approximate method based on the Central Limit Theorem (Ermakov and Mikhailov 1976).

Algorithm Groundwork

Choose a positive number n. A larger n produces a more accurate distribution, but the algorithm is slower. $n = 12$ has been recommended in the literature for applications where the tails of the distribution are not important.

Algorithm

1. Generate $\gamma_1, \gamma_2 \cdots \gamma_n$.
2. Calculate and output ξ:

$$\xi = \sqrt{\frac{12}{n}} \sum_{i=1}^{n} \left(\gamma_i - \frac{1}{2} \right). \tag{2.58}$$

By applying the technique we used in Eq. (2.57), it is easy to show that $\mathrm{Var}(\xi) = 1$ and $\mathrm{E}(\xi) = 0$.

Method 2

This method is based on the transformation of random variables in two dimensions. Instead of one random number, we sample two independent variables, ξ_1 and ξ_2, each having distribution $N(0, 1)$. The joint probability distribution of the two variables is:

$$f(x, y) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{x^2}{2} \right) \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{y^2}{2} \right). \tag{2.59}$$

Next, we introduce new variables, ρ and ϕ, defined as

$$x = \rho \cos\phi; \quad y = \rho \sin\phi, \tag{2.60}$$

and transform the joint distribution accordingly

$$f(x, y)\,\mathrm{d}x\mathrm{d}y = \frac{1}{2\pi} \exp\left(-\frac{\rho^2}{2} \right) \rho\,\mathrm{d}\rho\,\mathrm{d}\phi = \frac{\mathrm{d}\phi}{2\pi} \exp\left(-\frac{\rho^2}{2} \right) \mathrm{d}\left(\frac{\rho^2}{2} \right). \tag{2.61}$$

This result means that the joint distribution of ϕ and $\rho^2/2$ is a product of a uniform distribution and an exponential distribution:

$$f\left(\frac{\rho^2}{2}, \phi \right) = \frac{1}{2\pi} \exp\left(-\frac{\rho^2}{2} \right). \tag{2.62}$$

Algorithm Overview

1. Sample ϕ uniformly in the interval 0 to 2π.
2. Sample $\rho^2/2$ from the exponential distribution.
3. Transform the variables from $\rho^2/2$ and ϕ back to x and y.

Algorithm

1. Generate γ.
2. $\phi = 2\pi\gamma$.
3. Generate another γ.
4. $\eta = -\ln\gamma$, the CDF inversion method for the exponential distribution, where $\eta = \rho^2/2$.
5. $\rho = \sqrt{2\eta}$.
6. $\xi_1 = \rho\cos\phi$.
7. $\xi_2 = \rho\sin\phi$.

The downside to the above algorithm is having to calculate the sine and the cosine. This can be avoided. Let us consider a geometric interpretation of the algorithm, as shown in Fig. 2.6:

1. Sample from the distribution, isotropic in two dimensions, a direction that is defined by angle ϕ. This corresponds to steps 1 and 2 in the above algorithm.
2. From the origin, draw a straight line in that direction. It is shown in blue in Fig. 2.6.
3. Sample ρ, and place a point on the line, at distance ρ from the origin (blue dot in Fig. 2.6). This corresponds to steps 3–5 in the algorithm.
4. Find the projections ξ_1 and ξ_2 of that point on x and y axes, that is, the x and y coordinates of the dot. This is accomplished in steps 6 and 7.

Note that we do not actually need to know the angle ϕ. We only used it to find the x and y projections of the blue dot. Furthermore, there is another way of sampling an isotropic direction in two dimensions. Namely, if a point has a uniform distribution within a circle, then the line connecting the point and the origin is isotropically distributed. Equation (2.65) derived later in this chapter proves this statement.

Hence, we have **Method 3** (Fig. 2.7) for sampling the normal distribution $N(0, 1)$.

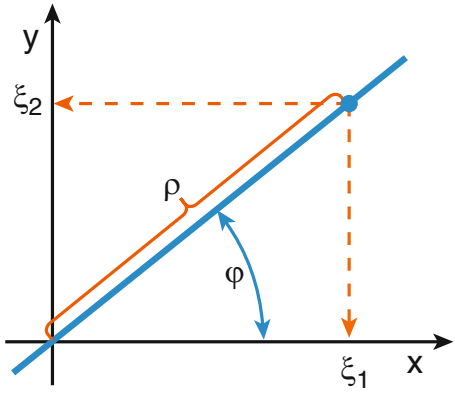

Fig. 2.6 Sampling a normal distribution. Method 2. Geometric interpretation of the algorithm

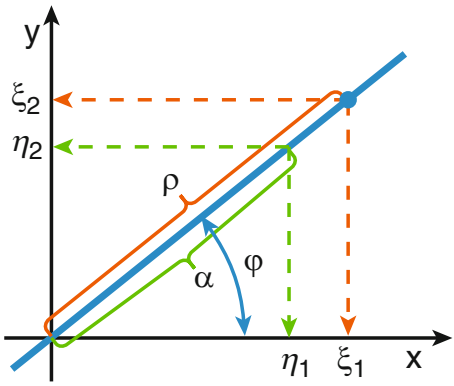

Fig. 2.7 Sampling normal distribution. Method 3

Algorithm

1. Generate γ_1 and γ_2.
2. Sample a point with coordinates (η_1, η_2) within a 2-by-2 square centered at the origin: $\eta_1 = 2\gamma_1 - 1$, $\eta_2 = 2\gamma_2 - 1$.
3. Calculate $\alpha^2 = \eta_1^2 + \eta_2^2$.
4. If $\alpha^2 > 1$, then point (η_1, η_2) is outside the unit circle. Try again, go to step 1.
5. Else, the point is within the circle.
6. Generate γ_3.
7. Calculate $\rho = \sqrt{-2 \ln \gamma_3}$ (here we sample $\rho^2/2$ from the exponential distribution).
8. Find x and y projections of the blue dot: $\xi_1 = \eta_1 \rho/\alpha$, $\xi_2 = \eta_2 \rho/\alpha$.

2.11 Random Points and Directions

Placing a Random Point Uniformly Within a Circle

The circle radius is R. It is centered at the origin of the coordinate system.

Method 1: Rejection

Algorithm

1. Generate γ_1 and γ_2.
2. Place a point within a $2R$ by $2R$ square: $\xi_1 = R(2\gamma_1 - 1)$, $\xi_2 = R(2\gamma_2 - 1)$. Here R is the radius of the circle.
3. If $\xi_1^2 + \xi_2^2 > R^2$, then the point is outside the circle; go to step 1.
4. Else, output coordinates (ξ_1, ξ_2) of the point.

Method 2: CDF Inversion

Algorithm Groundwork

The uniform distribution of a point within a circle in terms of x and y is a constant within the circle:

$$f(x,y) = \frac{1}{\pi R^2}; \quad \text{if } x^2 + y^2 < R^2. \tag{2.63}$$

The same distribution in polar coordinates is independent of ϕ and is linear in ρ:

$$f(\rho,\phi) = \frac{\rho}{\pi R^2}; \quad \text{if } \rho < R. \tag{2.64}$$

We have left out the derivation of Eq. (2.64) because it is very similar to the derivation of Eq. (2.62). This result means that ϕ and ρ are independent. Their respective distributions are

$$f(\phi) = \frac{1}{2\pi}; \quad \text{and } f(\rho) = \frac{2\rho}{R^2}. \tag{2.65}$$

Then, the algorithm states

1. Generate γ.
2. $\phi = 2\pi\gamma$.
3. Generate another γ.
4. $\rho = R\sqrt{\gamma}$ (the CDF inversion method for $f(\rho)$).
5. $\xi_1 = \rho\cos\phi$, $\xi_2 = \rho\sin\phi$ (x and y coordinates of the point).

Placing a Random Point Uniformly Within a Sphere

The sphere radius is R. It is centered at the origin of the coordinate system.

Method 1: Rejection

This algorithm is very similar to placing a point within a circle.

Method 2: The CDF Inversion Method

Algorithm Groundwork

In terms of x, y, and z, the uniform distribution within a sphere is a constant:

$$f(x,y,z) = \frac{1}{(4\pi/3)\,R^3}; \quad \text{if } x^2+y^2+z^2 < R^2. \tag{2.66}$$

In spherical coordinates, the same distribution can be written in the following form:

$$f(r,\theta,\phi) = \frac{3r^2}{R^3}\,\frac{\sin\theta}{2}\,\frac{1}{2\pi}; \quad \text{if } r < R. \tag{2.67}$$

Here, we again used the rule for transforming random variables, this time from three-dimensional Cartesian coordinates to spherical coordinates. This result means that r, θ, and ϕ are independent, with their respective distributions:

$$f(r) = \frac{3r^2}{R^3}; \quad f(\theta) = \frac{\sin\theta}{2}; \quad f(\phi) = \frac{1}{2\pi}. \tag{2.68}$$

Distributions $f(r)$ and $f(\phi)$ are easy to sample. Instead of θ, it is much easier to sample $\mu = \cos\theta$, where $\mu \in [-1,\ 1]$. Using Eq. (2.24), we find that

$$f(\mu) = \frac{1}{2}\,. \tag{2.69}$$

In other words, μ is distributed uniformly in $[-1, 1]$.

Algorithm

1. Generate γ.
2. $r = R\gamma^{1/3}$ (CDF inversion method for $f(r)$).
3. Generate another γ.
4. $\mu = 2\gamma - 1$ (μ is sampled uniformly in $[-1, 1]$).
5. Generate yet another γ.
6. $\phi = 2\pi\gamma$.
7. $\xi_1 = r\sin\theta\cos\phi$; $\xi_2 = r\sin\theta\sin\phi$; $\xi_3 = r\cos\theta$; where $\cos\theta = \mu$, $\sin\theta = \sqrt{1-\mu^2}$, and ξ_1, ξ_2, ξ_3 are the x, y, and z coordinates of the point.

Sampling a Direction from Isotropic Distribution

A unit vector $\vec{\Omega} = \{\Omega_x, \Omega_y, \Omega_z\}$ defines a direction in three dimensions. "Isotropic" means that, for any solid angle $\mathrm{d}\omega$,

$$P\{\vec{\Omega} \in \mathrm{d}\omega\} = \frac{\mathrm{d}\omega}{4\pi}. \tag{2.70}$$

A solid angle is an area on a unit sphere ($r = 1$), which in spherical coordinates, is

$$d\omega = \sin\theta d\theta d\phi = d\mu d\phi. \tag{2.71}$$

This means that μ and ϕ are independent and uniformly distributed, in $[-1, 1]$ and $[0, 2\pi)$, respectively. Then, to sample $\vec{\Omega}$ isotropically, the following obvious algorithm can be used:

Algorithm

1. Generate γ.
2. $\mu = 2\gamma - 1$.
3. Generate γ.
4. $\phi = 2\pi\gamma$.
5. Calculate the components of vector $\vec{\Omega}$: $\Omega_x = \sqrt{1-\mu^2}\cos\phi$; $\Omega_y = \sqrt{1-\mu^2}\sin\phi$; $\Omega_z = \mu$.

Placing a Random Point Uniformly on a Spherical Surface

The sphere radius is R. It is centered at the origin of the coordinate system.

Algorithm

1. Sample an isotropic vector $\vec{\Omega} = \{\Omega_x, \Omega_y, \Omega_z\}$, as described above.
2. Calculate x,y, and z coordinates of the point: $\xi_1 = \Omega_x R$; $\xi_2 = \Omega_y R$; $\xi_3 = \Omega_z R$; where R is the radius of the sphere.

2.12 Sampling a Joint Distribution

Problem

Sample two random numbers that are not statistically independent.

Example

See Fig. 2.8. When an incident particle ionizes a molecule, a secondary electron is ejected. Its energy and ejection angle (θ) are random variables. They are not independent.

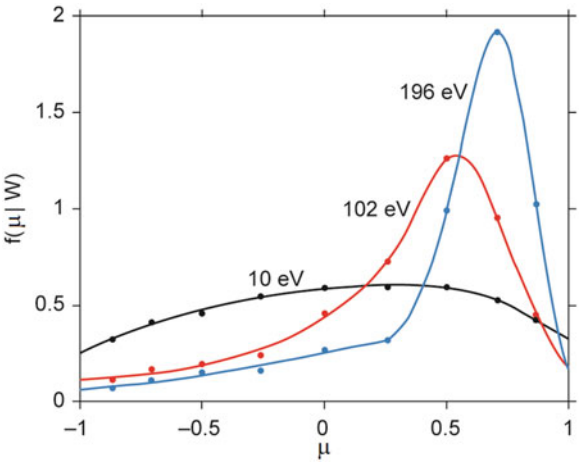

Fig. 2.8 Angular distributions of electrons ejected with energies 10, 102, and 196 eV from a water molecule by 500 eV electrons. The probability distributions $f(\mu|w)$ were derived from the experimental doubly differential cross sections reported by Opal et al. (1972) (*solid circles*). Here μ is the cosine of the scattering angle θ, and w is the energy of the ejected electron. The *lines* are drawn only to guide the eye

Algorithm Groundwork

The joint distribution of two random variables ξ and η is given by $f_{\xi,\eta}(x, y)$. Using the formula for conditional probability, we can rewrite it as

$$f_{\xi,\eta}(x, y) = f_{\xi}(x) f_{\eta}(y|\xi = x), \tag{2.72}$$

where $f_{\xi}(x)$ is the distribution of ξ. It can be found by integrating the joint distribution

$$f_{\xi}(x) = \int_{-\infty}^{+\infty} f_{\xi,\eta}(x, y)\mathrm{d}y. \tag{2.73}$$

Distribution $f_{\eta}(y|\xi = x)$ is the conditional distribution of η for a given value of ξ:

$$f_{\eta}(y|\xi = x) = \frac{f_{\xi,\eta}(x, y)}{f_{\xi}(x)}. \tag{2.74}$$

Algorithm

Equation (2.72) suggests the following algorithm:

1. Sample ξ from distribution $f_{\xi}(x)$.
2. Sample η from the conditional distribution $f_{\eta}(y|\xi = x)$.

When simulating an ionization event, usually, the energy of the ejected electron is sampled first, then its ejection angle is sampled.

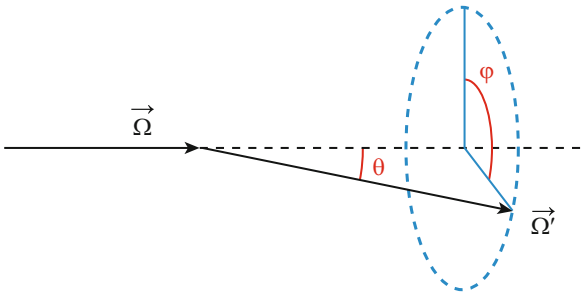

Fig. 2.9 Particle-scattering event

2.13 Simulating a Particle-Scattering Event

The parameters characterizing a particle scattering event are shown in Fig. 2.9. In this figure, $\vec{\Omega}$ is the particle direction before the collision, $\vec{\Omega}'$ is its direction after the collision, θ is the scattering angle (polar angle), ϕ is the azimuthal angle.

Algorithm

1. Sample the particle energy after scattering, E'. Its distribution is given by the singly differential cross section.
2. Sample the scattering angle θ; its distribution is derived from the doubly differential cross section for a given E'.
3. Sample the azimuthal angle ϕ, usually uniformly, in $[0, 2\pi)$.
4. Given $\vec{\Omega}$, θ and ϕ, calculate the particle direction after collision, $\vec{\Omega}' = \{\Omega'_x, \Omega'_y, \Omega'_z\}$.

Step 4 is simple if $\vec{\Omega}$ is parallel to one of the axes, for example the z-axis, that is, $\vec{\Omega} = \{0, 0, 1\}$. In that case,

$$\Omega'_x = \sin\theta \cos\phi. \tag{2.75}$$

$$\Omega'_y = \sin\theta \sin\phi. \tag{2.76}$$

$$\Omega'_z = \cos\theta. \tag{2.77}$$

For an arbitrary $\vec{\Omega} = \{\Omega_x, \Omega_y, \Omega_z\}$, the formulae are:

$$\Omega'_x = \Omega_x \cos\theta - \frac{\sin\theta}{\sqrt{1-\Omega_z^2}} \left(\Omega_x \Omega_z \cos\phi + \Omega_y \sin\phi\right); \tag{2.78}$$

$$\Omega'_y = \Omega_y \cos\theta - \frac{\sin\theta}{\sqrt{1-\Omega_z^2}} \left(\Omega_y \Omega_z \cos\phi - \Omega_x \sin\phi\right); \tag{2.79}$$

$$\Omega'_z = \Omega_z \cos\theta + \sqrt{1-\Omega_z^2}\sin\theta\cos\phi. \tag{2.80}$$

A derivation of the above equations can be found, for example, in Haghihat (2015).

Sampling ϕ and then calculating $\sin\phi$ and $\cos\phi$ is inefficient. Instead, $\sin\phi$ and $\cos\phi$ can be sampled directly, without having to sample ϕ first. In Sect. 2.10, we have already introduced this technique for sampling normal distributions using Method 3.

Here is the algorithm for sampling $\cos\phi$ and $\sin\phi$:

Algorithm

1. Sample γ.
2. $\eta_1 = 1 - 2\gamma$.
3. Sample another γ.
4. $\eta_2 = 1 - 2\gamma$.
5. $\alpha = \eta_1^2 + \eta_2^2$.
6. If $\alpha > 1$, go to step 1.
7. Output $\cos\phi = \eta_1/\sqrt{\alpha}$, and $\sin\phi = \eta_2/\sqrt{\alpha}$.

The probability of rejection at step 6 ($\alpha > 1$) is $1 - \pi/4$.

2.14 Algorithm Testing

In this section we discuss methods for testing software developed for generating random numbers with a given distribution $f(x)$. The hypothesis we put to the test is that the software indeed generates random numbers with distribution $f(x)$.

2.14.1 Histograms

The first method is qualitative. The code is run to generate a large sample $\{\xi_1, \xi_2, \ldots, \xi_N\}$, $N \gg 1$. These data are represented graphically as a histogram. On the same graph, for comparison, distribution $f(x)$ is plotted. If the code works correctly the line representing $f(x)$ passes through the centers of the histogram bars at their tops, as can be seen in Fig. 2.10. For this test, the histogram must be normalized, so that the total area under the histogram is equal to 1, Thus

$$\sum_i h_i \Delta x_i = 1, \tag{2.81}$$

where Δx_i is the histogram bin width and h_i is its height.

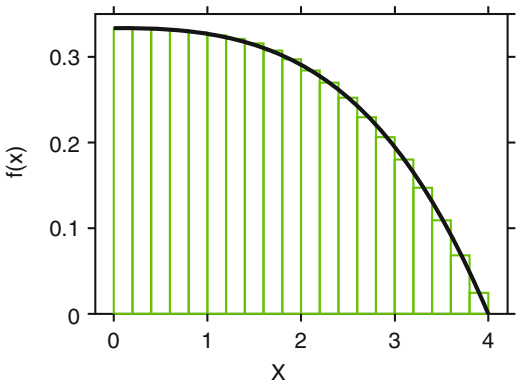

Fig. 2.10 A histogram compared with a probability density function $f(x)$

Histogram bins should be defined so that they cover all possible values of the random value. The check

$$\sum_i n_i = N, \tag{2.82}$$

should be performed, where n_i is the number of random values that fell within the i-th bin, and N is the sample size. The height of the bar, h_i, is calculated as follows:

$$h_i = \frac{n_i}{N\Delta x_i}. \tag{2.83}$$

Then, the histogram is correctly normalized:

$$\sum_i h_i \Delta x_i = \sum_i \left(\frac{n_i}{N\Delta x_i} \right) \Delta x_i = \frac{1}{N} \sum_i n_i = 1. \tag{2.84}$$

If distribution $f(x)$ has a singularity, then a separate very small bin around the singular point should be included. And, if $f(x)$ has a discontinuity, a bin boundary should be placed exactly at the point of discontinuity.

Example

Annihilation photons have an initial energy close to the rest energy of an electron: $E_\gamma \approx m_e c^2 \approx 511\,\text{keV}$. In a photon beam produced by a medical accelerator annihilation photons present only a small fraction of the total fluence. However, in the histogram estimate of the fluence spectrum, they produce a distinct spike, shown in Fig. 2.11, provided that a narrow bin is designed around 511 keV. The spectrum shown in Fig. 2.11 was generated with the Monte Carlo method using an accelerator model developed by Cho et al. (2005).

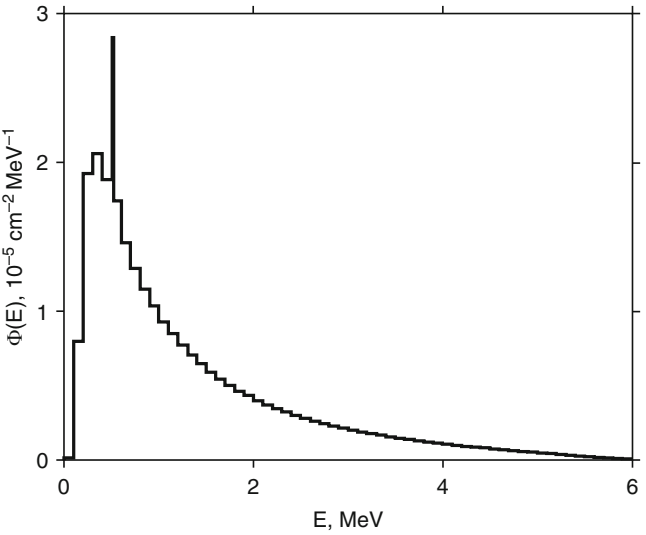

Fig. 2.11 Fluence spectrum of a 6-MV photon beam from a medical accelerator. The spike at 511 keV is due to annihilation photons. The bin width is 7.5 keV at 511 keV and 50 keV elsewhere

2.14.2 The χ^2 Test

The χ^2 test is a quantitative pass/fail test. To perform it, follow this step-by-step procedure:

1. Start by dividing the x-axis into intervals: $1, 2, \ldots, k$, as shown in Fig. 2.12. Intervals can be infinite. For example, Bin # 6 spans from 2 to infinity.
2. Assuming that the distribution of the random variable ξ is $f(x)$, calculate probabilities p_i of ξ falling within each interval. This requires calculating the integrals

$$p_i = \int_{a_i}^{b_i} f(x)\,\mathrm{d}x; \quad i = 1, 2, \ldots, k, \tag{2.85}$$

   where a_i and b_i are the lower and upper boundaries of interval i, respectively, and values of p_i are called the predicted, or expected, probabilities.
3. Then, run the code to generate a large sample: $\xi_1, \xi_2 \ldots, \xi_N$, $N \gg 1$. For each interval, find n_i, the number of random numbers in the sample that fell within interval i.

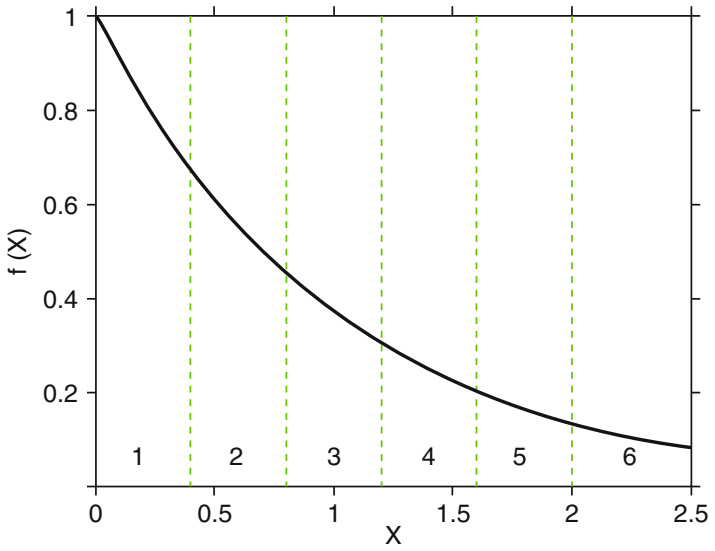

Fig. 2.12 χ^2 test

4. Calculate the "observed" probabilities

$$\pi_i = \frac{n_i}{N}; \quad i = 1, 2, \ldots, k \tag{2.86}$$

and "the observed" χ^2

$$\chi^2 \text{ (observed)} = \sum_{i=1}^{k} \frac{(\pi_i - p_i)^2}{p_i}. \tag{2.87}$$

It should be possible to generate a very large sample and thus ensure that all π_i are approximately normally distributed. That requires that each n_i is much greater than 1. Then, the above sum is χ^2-distributed (see Appendix A). If, additionally, the number of intervals k is sufficiently large, it can be assumed that the above sum is normally distributed.

5. Using the χ^2 distribution or, if applicable, the normal distribution, calculate the probability of observing a χ^2 that is equal or greater than χ^2 (observed), using

$$P\{\chi^2 \geq \chi^2 \text{ (observed)}\} = \int_{\chi^2(\text{observed})}^{\infty} p_{\chi^2}(x)\, \mathrm{d}x, \tag{2.88}$$

where p_{χ^2} is the appropriate distribution of χ^2 (probability density).

6. If the above probability is small, for example less than 0.05 or 0.001, the hypothesis is rejected. This means that if we assume that the code works correctly, then χ^2 (observed) is unrealistically large and there is a good chance that the code samples some other distribution, not the desired distribution $f(x)$.

2.14.3 The Likelihood Ratio Test

We present the likelihood ratio test in a qualitative form, which suffices for our purposes, owing to the sensitivity of the test to algorithm errors.

Note 1. Distribution $f(x)$ and all its parameters are known exactly. We will write it $f(x|\theta_0)$, where θ_0 are the exact parameters.

Note 2. $f(x|\theta)$ denotes exactly the same distribution (the same functional form), except it was calculated with the exact parameters (θ_0) replaced by their maximum likelihood estimates θ.

To perform the test:

1. Run the code, generating a large sample $\{\xi_1, \xi_2, \ldots, \xi_N\}$, $N \gg 1$.
2. Use the sample to estimate the parameters of the distribution $f(x)$.
3. Calculate the likelihood ratio vector:

$$\lambda_i = \prod_{j=1}^{i} \frac{f\left(\xi_j|\theta_0\right)}{f\left(\xi_j|\theta\right)}; \quad i = 1, 2, \ldots, N. \tag{2.89}$$

4. Plot λ_i as a function of i. If the code works correctly, λ_i will fluctuate around 1. If it does not, λ_i will be diverging.

Example

Testing a code that samples the normal distribution $N(1, 1)$, Fig. 2.13.

In Fig. 2.13, the red line shows the likelihood ratio for a code intended to sample the normal distribution $N(1, 1)$; it does so almost correctly, except for a 1 % error in the standard deviation. The blue line shows a similar result, but this time with a 1 % error in the mean of the distribution. The black line shows the result for the code working correctly.

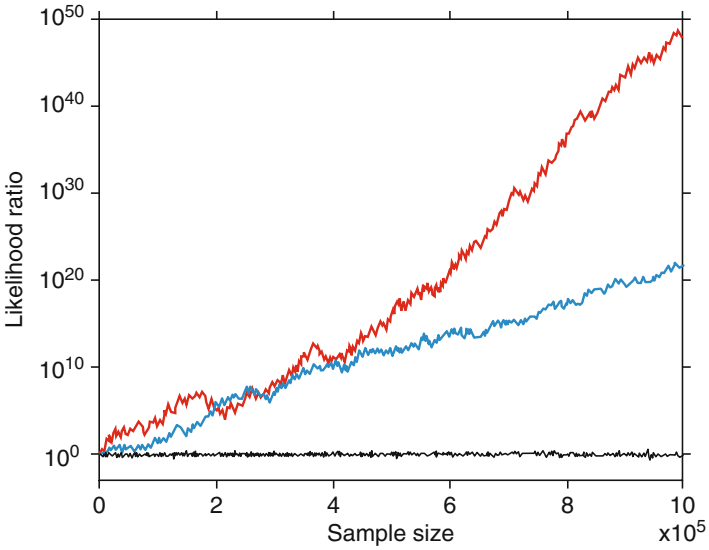

Fig. 2.13 The likelihood ratio test for a code sampling normal distribution $N(1, 1)$. The *black line* shows the likelihood ratio for a correct code, the *blue line* indicates the code with a 1 % error in the mean, and the *red line* represents a 1 % error in the standard deviation

References

Attix, F.H.: Introduction to Radiological Physics and Radiation Dosimetry. Wiley, New York (2004)

Butcher, J.C., Messel, H.: Electron number distribution in electron-photon showers in air and aluminum absorbers. Nucl. Phys. **20**(1), 15–128 (1960)

Cho, S.H., Vassiliev, O.N., Lee, S., Ibbott, G.S., Mohan, R.: Reference photon dosimetry data and reference phase space data for the 6 MV photon beam from Varian Clinac 2100 series linear accelerators. Med. Phys. **32**(1), 137–148 (2005)

Ermakov, S.M., Mikhailov, G.A.: Course of Statistical Modelling. Nauka, Moscow (1976)

Fishman, G.S.: Monte Carlo: Concepts, Algorithms, and Applications. Springer, New York (1996)

Ford, R.L., Nelson, W.R.: The EGS Code System: computer programs for the Monte Carlo simulation of electromagnetic cascade showers (Version 3). Stanford Linear Accelerator Center, Report 210 (1978)

Haghihat, A.: Monte Carlo Methods for Particle Transport. CRC, Boca Raton, FL (2015)

Lehmer, D.H.: Mathematical methods in large-scale computing units. Ann. Comput. Lab. Harv. Univ. **26**(1), 141–146 (1951)

Opal, C.B., Beaty, E.C., Peterson, W.K.: Tables of secondary-electron-production cross sections. At. Data **4**, 209–253 (1972)

Chapter 3
The Boltzmann Equation

3.1 Definitions

Phase Coordinates

The phase coordinate of a particle is its position and momentum: $(\vec{r}, \vec{p}\,)$. It is a six-dimensional variable, $\left(x, y, z, p_x, p_y, p_z\right)$, a point in a six-dimensional phase space. In this book, instead of momentum $\vec{p}$ we will mostly use two variables: particle energy E and a unit vector: $\vec{\Omega} = \vec{p}/|\vec{p}\,|$, that defines the direction of particle travel. This increases the number of variables to seven: $\left(x, y, z, \Omega_x, \Omega_y, \Omega_z, E\right)$. However, the dimensionality remains six, that is, only six variables are independent because $\vec{\Omega}$ is a unit vector:

$$\Omega_x^2 + \Omega_y^2 + \Omega_z^2 = 1. \tag{3.1}$$

We will use the following notations for phase coordinates: $(\vec{r}, \vec{\Omega}, E)$, or simply x. That is, $f(x)$ is the same as $f(\vec{r}, \vec{\Omega}, E)$. Additionally, the particle state can be characterized by internal degrees of freedom, which usually are discrete variables, for example, the charge state for charged particles (C^+ ion versus C^{2+}), polarization for photons, and, most commonly, particle type (γ, e^-, e^+, p, n, ...).

Source Function

The source function $S(x)$ fully describes the radiation source: the spatial $(\vec{r})$, angular $(\vec{\Omega})$, and energy (E) distributions of particles emitted by the source. In the Monte Carlo method, it is convenient to normalize the source function per one particle:

O.N. Vassiliev, *Monte Carlo Methods for Radiation Transport*, Biological and Medical Physics, Biomedical Engineering, DOI 10.1007/978-3-319-44141-2_3

$$\int S(x)\,\mathrm{d}x = 1, \tag{3.2}$$

so that $S(x)$ could be interpreted as the probability distribution function of x.

Total Cross Section

The concept of cross sections can be understood in a classical model of interaction: an atom is a sphere, and a particle interacts with the atom if it enters the sphere. Then, the probability of interaction is proportional to the cross-sectional area of the sphere, πR^2. Accordingly, the units for cross section are cm^2. Cross sections measured in cm^2 are sometimes referred to as *microscopic*. In this book, however, we will never use these units. Instead, we will use the so-called *macroscopic* cross sections defined as follows:

σ (macroscopic) = microscopic cross section $\times$ number of atoms per cm^3.

In that case, the units for σ are $\mathrm{cm}^2 \times \mathrm{cm}^{-3} = 1/\mathrm{cm}$. The following are examples that clarify the meaning of the macroscopic cross section σ:

1. The probability of a particle interacting as it travels distance $\mathrm{d}l$ is:

$$P(\mathrm{d}l) = \sigma \mathrm{d}l. \tag{3.3}$$

   This equation is, in fact, the definition of the macroscopic cross section.
2. The probability of a particle traveling distance l without interactions is:

$$P\{n = 0\} = \exp(-\sigma l), \tag{3.4}$$

   where n is the number of interactions.

Proof of Eq. (3.4). Let $P(l)$ be the probability of a particle traveling distance l without interactions. Then the probability that it travels without an interaction distance $l + \mathrm{d}l$ is a product of two probabilities: $P(l)$ and the probability that it travels distance $\mathrm{d}l$ without interactions:

$$P(l + \mathrm{d}l) = P(l)\,(1 - \sigma \mathrm{d}l)\,. \tag{3.5}$$

This leads us to a simple differential equation. Solving it with the boundary condition $P(0) = 1$ produces Eq. (3.4). □

The distance l a particle travels between two consecutive interactions is called the free path. Note that Eq. (3.4) is valid only if the particle travels in a homogeneous material. Generally, cross sections may change along the particle track, that is, $\sigma = \sigma(\vec{r}\,)$. A solution for this general case is given below, in Eq. (3.11).

3. The probability density of free path is:

$$f(l) = \sigma \exp(-\sigma l). \tag{3.6}$$

Proof of Eq. (3.6).

$$f(l) = \frac{\mathrm{d}}{\mathrm{d}l} F(l) = \frac{\mathrm{d}}{\mathrm{d}l} [1 - P(l)] = \sigma \exp(-\sigma l), \tag{3.7}$$

where $F(l)$ is the cumulative distribution function. □

4. The mean free path:

$$\langle l \rangle = \int_0^\infty l \cdot f(l)\, \mathrm{d}l = \frac{1}{\sigma}. \tag{3.8}$$

In other words, σ is the probability of an interaction per unit path length. If this means the probability of any interaction, then σ is called the total cross section and is sometimes written as σ_{tot}. If, on the other hand, we are concerned with a particular type of interaction, then σ is called the partial cross section, or, if the process in question is, for example, absorption, it is called the absorption cross section. The total cross section is the sum of all partial cross sections.

Optical Distance

The definition of optical distance implies that between two consecutive interaction points the particle travels along a straight line. The equation of a straight line that is parallel to $\vec{\Omega}$ and passes through point $\vec{r}_0$, Fig. 3.1, can be written as follows:

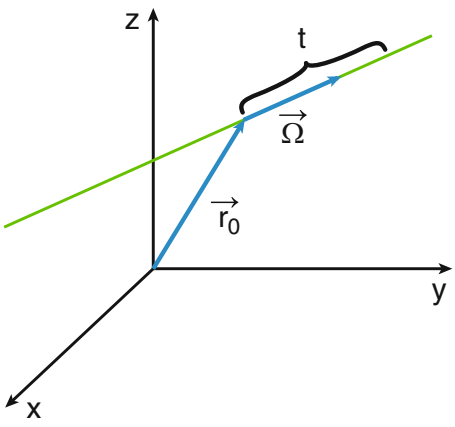

Fig. 3.1 To the definition of optical distance

$$\vec{r}\,(t) = \vec{r}_0 + \vec{\Omega} t; \quad t \in (-\infty, +\infty)\,. \tag{3.9}$$

Then, the optical distance between points $\vec{r}_1$ and $\vec{r}_2 = \vec{r}_1 + \vec{\Omega} l$ is defined as an integral along a straight line connecting the two points:

$$\tau\left(\vec{r}_1, \vec{r}_2\right) = \int_0^l \sigma\left(\vec{r}_1 + \vec{\Omega} t\right) \mathrm{d}t. \tag{3.10}$$

In a heterogeneous medium, Eq. (3.4) takes the form:

$$P\{n = 0\} = \exp\left[-\tau\left(\vec{r}_1, \vec{r}_2\right)\right]. \tag{3.11}$$

It can be derived by integrating Eq. (3.5) without assuming a constant σ.

Differential Cross Sections

Speaking generally, differential cross sections define the probability of an interaction with a certain outcome. For example, for the doubly differential cross section the following is the probability that a particle with parameters $(\vec{\Omega}, E)$ undergoes an interaction as it travels distance $\mathrm{d}l$ and its direction after the interaction is within the solid angle $\mathrm{d}\vec{\Omega}'$ and its energy is within $\mathrm{d}E'$:

$$\sigma_s\left(\vec{r}, \vec{\Omega} \to \vec{\Omega}', E \to E'\right) \mathrm{d}l \mathrm{d}\vec{\Omega}' \mathrm{d}E'. \tag{3.12}$$

The total cross section is the integral of the differential cross section:

$$\sigma_s\left(\vec{r}, E\right) = \int \mathrm{d}\vec{\Omega}' \int \mathrm{d}E' \sigma_s\left(\vec{r}, \vec{\Omega} \to \vec{\Omega}', E \to E'\right). \tag{3.13}$$

In the above equation, the cross section σ_s on the left-hand side is independent of the particle direction $\vec{\Omega}$ before the interaction. It is usually assumed that the material properties are isotropic and, therefore, the interaction probability does not depend on the direction of particle travel. If the number of particles after the collision is more than one, including when it is random, then Eq. (3.12) instead of the probability defines the average number of particles with parameters within $\mathrm{d}\vec{\Omega}'$ and $\mathrm{d}E'$.

Transport Cross Section

It is also often assumed that the differential scattering cross section is a function of the cosine of scattering angle $\cos\theta_0 \equiv \mu_0 = (\vec{\Omega} \cdot \vec{\Omega}')$. Then, the transport cross section is defined as follows:

$$\sigma_{\text{tr}}\left(\vec{r},E\right)\equiv\int \mathrm{d}\vec{\Omega}'\sigma_s\left(\vec{r},\vec{\Omega}\cdot\vec{\Omega}',E\right)\left(1-\vec{\Omega}\cdot\vec{\Omega}'\right), \tag{3.14}$$

where

$$\sigma_s\left(\vec{r},\vec{\Omega}\cdot\vec{\Omega}',E\right)=\int \mathrm{d}E'\sigma_s\left(\vec{r},\vec{\Omega}\cdot\vec{\Omega}',E\rightarrow E'\right). \tag{3.15}$$

Transport cross section is an important quantity. For example, in neutron transport, if the diffusion approximation is applicable, the transport cross section defines the diffusion coefficient.

Fluence: Definition 1

Fluence $\Phi(\vec{r},\vec{\Omega},E)$ is the average number of particles with energy E crossing a unit area of a surface located at point $\vec{r}$ and oriented normally to $\vec{\Omega}$.

Figure 3.2 illustrates this definition. Here surface dA is centered at point $\vec{r}$ and oriented normally to $\vec{\Omega}$. Then, according to the definition the number of particles dN crossing dA is

$$\mathrm{d}N\left(\vec{r},\vec{\Omega},E\right)=\Phi\left(\vec{r},\vec{\Omega},E\right)\mathrm{d}A\,. \tag{3.16}$$

Here dN can be either the number of particles crossing the surface per source particle, or per unit time. This depends on how the source function $S(x)$ is normalized, per one particle, or per one particle per unit time. If the surface is not normal to $\vec{\Omega}$, then fluence is multiplied by the cosine of the angle between $\vec{\Omega}$ and normal vector to the surface, $\vec{n}$, Fig. 3.3:

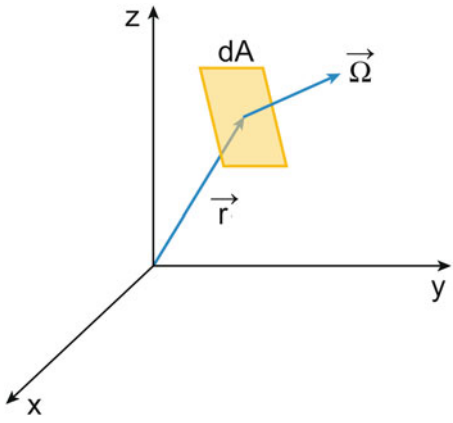

Fig. 3.2 To definition 1 of fluence

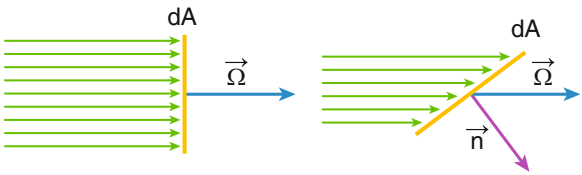

Fig. 3.3 To definition 1 of fluence

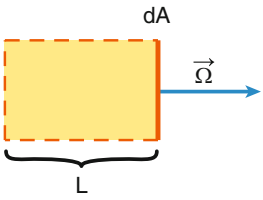

Fig. 3.4 To definition 2 of fluence

$$\mathrm{d}N\left(\vec{r},\vec{\Omega},E\right)=\Phi\left(\vec{r},\vec{\Omega},E\right)|\vec{\Omega}\cdot\vec{n}|\mathrm{d}A. \tag{3.17}$$

Here we take the absolute value of the cosine, because fluence is nonnegative, and because in doing so we eliminate the ambiguity associated with choosing one of the two possible direction of the normal.

There is an alternative way to define fluence. Let $n(\vec{r},\vec{\Omega},E)$ be the number density, i.e., the number of particles with parameters $(\vec{\Omega},E)$ per unit volume, and L be the distance traveled by a particle with parameters $(\vec{\Omega},E)$, Fig. 3.4. Then, only particles within the shaded volume will reach the surface area dA leading to:

$$\mathrm{d}N\left(\vec{r},\vec{\Omega},E\right)=n\left(\vec{r},\vec{\Omega},E\right)L\mathrm{d}A\,, \tag{3.18}$$

and, comparing with Eq. (3.16),

$$\Phi\left(\vec{r},\vec{\Omega},E\right)=n\left(\vec{r},\vec{\Omega},E\right)L\,. \tag{3.19}$$

Note, all the quantities in the above two equations, dN, n and L, are average. It is important to understand that fluence characterizes the average properties of a radiation field. In the right-hand side of Eq. (3.19), we have the number of particles per volume times the distance one particle travels. This brings us to the next definition.

Fluence: Definition 2

Fluence $\Phi(x)$ is the total distance traveled by particles with phase coordinate x, per unit volume. Here total distance means it is a sum of distances traveled by all individual particles.

There is yet another definition of fluence. Let us now find the number of particles entering a small sphere. The sphere is so small that the fluence at all points within the sphere is approximately the same. For particles entering the sphere we have: $(\vec{\Omega} \cdot \vec{n}) < 0$, as can be seen in Fig. 3.6. Then, the number of particles entering the sphere is an integral over the surface of the sphere, but only over those parts where $(\vec{\Omega} \cdot \vec{n}) < 0$:

$$N\left(\vec{r}, \vec{\Omega}, E\right) = -\oint_{\left(\vec{\Omega}\cdot\vec{n}\right)<0} \Phi\left(\vec{r}, \vec{\Omega}, E\right)\left(\vec{\Omega} \cdot \vec{n}\right) \mathrm{d}A. \tag{3.20}$$

Given that fluence is approximately constant within the sphere, the integral can be calculated:

$$N\left(\vec{r}, \vec{\Omega}, E\right) = -\Phi\left(\vec{r}, \vec{\Omega}, E\right) \oint_{\left(\vec{\Omega}\cdot\vec{n}\right)<0} \left(\vec{\Omega} \cdot \vec{n}\right) \mathrm{d}A = \Phi\left(\vec{r}, \vec{\Omega}, E\right) \pi R^2, \tag{3.21}$$

where πR^2 is the cross-sectional area of the sphere. This is because $|\vec{\Omega} \cdot \vec{n}|\mathrm{d}A$ is the area projected by $\mathrm{d}A$ on a plane normal to $\vec{\Omega}$.

Up to this point, we have considered particles traveling in a direction $\vec{\Omega}$. The number of particles traveling in any direction and entering the sphere is an integral over all directions:

$$N\left(\vec{r}, E\right) = \int N\left(\vec{r}, \vec{\Omega}, E\right) \mathrm{d}\vec{\Omega} = \Phi\left(\vec{r}, E\right) \pi R^2, \tag{3.22}$$

where

$$\Phi\left(\vec{r}, E\right) = \int \Phi\left(\vec{r}, \vec{\Omega}, E\right) \mathrm{d}\vec{\Omega} \tag{3.23}$$

is the fluence spectrum. Still, we are counting only particles with energy E. To include all energies, we integrate over E:

$$N\left(\vec{r}, \text{all energies, all directions}\right) = \Phi\left(\vec{r}\,\right) \pi R^2, \tag{3.24}$$

where

$$\Phi\left(\vec{r}\,\right) = \int \Phi\left(\vec{r}, E\right) \mathrm{d}E \tag{3.25}$$

is the total fluence. Thus, we have one more definition, one that is well known. It is exactly how fluence is defined in ICRU 60 (1998).

Fluence: Definition 3

The **fluence**, Φ, is the quotient of dN by da, where dN is the number of particles incident on a sphere of cross-sectional area da, thus

$$\Phi = \frac{\mathrm{d}N}{\mathrm{d}a}. \tag{3.26}$$

Unit: m^{-2}.

The fluence defined above (Definition 3) is used rarely in this book. We will refer to this fluence as the total fluence. Throughout this book, "fluence" means fluence as a function of phase coordinates, e.g., $\Phi(\vec{r}, \vec{\Omega}, E)$. We also avoid using the term differential fluence, as it may suggest calculating derivatives, which is incorrect. Instead, we prefer such terms as fluence distribution, e.g., angular or in energy, or fluence energy spectrum.

Collision Density

Given that fluence is defined as the distance traveled by particles (definition 2) and σ is the probability of an interaction per unit path length, the product $\sigma\Phi$ is the average number of collisions. It is called the collision density and denoted F. More specifically:

$$F(x)\,\mathrm{d}x = \sigma\left(\vec{r}, E\right)\Phi(x)\,\mathrm{d}x \tag{3.27}$$

is the average number of collisions in phase space volume dx.

Detector Response Function

"Detector" is a mathematical concept that may or may not represent an actual physical device. The radiation field is characterized by fluence $\Phi(x)$, and quantities of interest ("observables") J are linear functionals of $\Phi(x)$:

$$J = \int \Phi(x)\,D(x)\,\mathrm{d}x, \tag{3.28}$$

where $D(x)$ is called the detector response function. The meaning of $D(x)$ can be inferred from definition 2 of fluence and the above integral: $D(x)$ is the contribution to an observable from a unit path length of a particle with a phase coordinate x. Quantity J is also referred to as the detector reading in field $\Phi(x)$. .

Example 1

The observable J is the ionization density, that is, the average number of ionizations in a volume V, per unit volume. In this case

$$D(x) = \begin{cases} \frac{1}{V}\sigma_i\left(\vec{r}, E\right), & \vec{r} \in V; \\ 0, & \vec{r} \notin V; \end{cases} \tag{3.29}$$

where σ_i is the ionization cross section and

$$J(V) = \frac{1}{V}\int_V \mathrm{d}\vec{r} \int \mathrm{d}\vec{\Omega} \int \mathrm{d}E\, \sigma_i\left(\vec{r}, E\right) \Phi\left(\vec{r}, \vec{\Omega}, E\right). \tag{3.30}$$

Example 2

The observable J is the absorbed dose at a point $\vec{r}$. Vassiliev et al. (2010) used the following expression for electron dose:

$$J\left(\vec{r}\right) \equiv \mathrm{Dose}\left(\vec{r}\right) = \int_0^\infty \mathrm{d}E \int_{4\pi} \mathrm{d}\vec{\Omega} \frac{\sigma^e_{\mathrm{ED}}\left(\vec{r}, E\right)}{\rho} \Phi^e\left(\vec{r}, \vec{\Omega}, E\right). \tag{3.31}$$

Here, σ_{ED} is called the energy deposition cross section, ρ is the mass density, and Φ^e is electron fluence. If $\beta\left(\vec{r}, E\right)$ is the electron stopping power, defined as energy lost by an electron per unit path length, and we can assume that energy lost is equal to energy deposited, then

$$\sigma_{\mathrm{ED}} = \beta\left(\vec{r}, E\right). \tag{3.32}$$

Hence, $D(x) = \beta/\rho$.

Adjoint Function

Alternatively, observables can be represented as the following functional:

$$J = \int S(x)\, \Phi^+(x)\, \mathrm{d}x, \tag{3.33}$$

where S is the source function and Φ^+ is called the adjoint function. By definition, which is consistent with Eq. (3.33), $\Phi^+(x)$ is the average contribution to an observable from a particle originating at a phase point x. $\Phi^+(x)$ is also called the importance function.

In this book, for brevity, we will often use the inner product notations for integrals over the phase space, for example:

$$J = (\Phi, D) = (S, \Phi^{+}). \tag{3.34}$$

3.2 Introduction to the Boltzmann Equation

Balance Equation in One Dimension

We start with a very simple case of a one-dimensional steady flow of particles, Fig. 3.5. Steady means that fluence at any given point does not change with time. Hence, we are concerned only with its spatial variations. For the number of particles in a small volume ΔV we can now write an obvious equation of balance:

$$\{\text{number of particles entering } \Delta V\} + \{\text{particles produced by a source within } \Delta V\} = \{\text{number of particles leaving } \Delta V\}.$$

In this model we have two reasons why the number of particles in ΔV can change: (1) particles enter and leave the volume through its surface, and (2) particles are produced by a source within ΔV. Using definition 1 of fluence the above balance equation can be written in terms of fluence and source function:

$$\Phi(x) A + S(x) \Delta x A = \Phi(x + \Delta x) A, \tag{3.35}$$

where A is the cross-sectional area of the beam. Dividing both sides of the equation by $A\Delta x$ and taking the limit $\Delta x \to 0$ brings us to a one-dimensional form of the balance equation:

$$\frac{\mathrm{d}\Phi(x)}{\mathrm{d}x} = S(x). \tag{3.36}$$

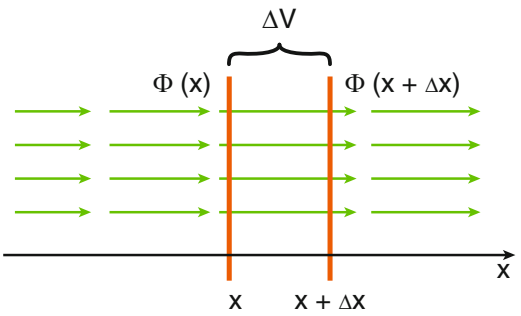

Fig. 3.5 Balance equation in one dimension

The derivative in the left-hand side of the equation is a one-dimensional form of the streaming operator. It accounts for particles entering and leaving the volume through its surface.

Balance Equation in Three Dimensions

This equation is very similar to the continuity equation in fluid mechanics. The difference is that the continuity equation expresses mass conservation and is written in terms of the fluid velocity and mass density. Here, however, we are concerned with conservation of the number of particles and write the balance equation in terms of fluence. We can start with exactly the same equation as in the one-dimensional case:

$$\{\text{number of particles entering } \Delta V\} + \{\text{particles produced by a source within } \Delta V\} = \{\text{number of particles leaving } \Delta V\}.$$

This time, however, to account for particles entering and leaving the volume, we need to integrate over its surface, $\Gamma(\Delta V)$. If $\vec{n}$ is an outward-pointing unit vector normal to the surface Γ, then to find the number of particles entering the volume we will integrate over the surface Γ only where $(\vec{\Omega} \cdot \vec{n}) < 0$, Fig. 3.6. Conversely, for particles leaving the volume we will integrate only where $(\vec{\Omega} \cdot \vec{n}) > 0$. Furthermore, now particles can in principle travel in any direction. We account for this by including the directional dependence of fluence. Again using definition 1 of fluence, we then arrive at an integral form of the balance equation:

$$-\oint_{\Gamma:\left(\vec{\Omega}\cdot\vec{n}\right)<0} \Phi\left(\vec{r},\vec{\Omega},E\right)\left(\vec{\Omega}\cdot\vec{n}\right)\mathrm{d}A + \int_{\Delta V} S\left(\vec{r},\vec{\Omega},E\right)\mathrm{d}V = \oint_{\Gamma:\left(\vec{\Omega}\cdot\vec{n}\right)>0} \Phi\left(\vec{r},\vec{\Omega},E\right)\left(\vec{\Omega}\cdot\vec{n}\right)\mathrm{d}A. \tag{3.37}$$

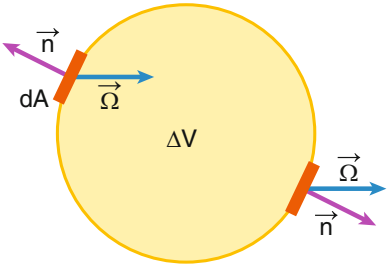

Fig. 3.6 Balance equation in three dimensions

These two surface integrals can be combined into one and then, using the divergence theorem, transformed into a volume integral:

$$\begin{aligned}
&\oint_{\Gamma:\left(\vec{\Omega}\cdot\vec{n}\right)<0} \Phi\left(\vec{r},\vec{\Omega},E\right)\left(\vec{\Omega}\cdot\vec{n}\right)\mathrm{d}A + \oint_{\Gamma:\left(\vec{\Omega}\cdot\vec{n}\right)>0} \Phi\left(\vec{r},\vec{\Omega},E\right)\left(\vec{\Omega}\cdot\vec{n}\right)\mathrm{d}A \\
&= \oint_{\Gamma} \Phi\left(\vec{r},\vec{\Omega},E\right)\left(\vec{\Omega}\cdot\vec{n}\right)\mathrm{d}A = \int_{\Delta V} \mathrm{div}\left[\vec{\Omega}\Phi\left(\vec{r},\vec{\Omega},E\right)\right]\mathrm{d}V \\
&= \int_{\Delta V} \left(\vec{\Omega}\cdot\vec{\nabla}\right)\Phi\left(\vec{r},\vec{\Omega},E\right)\mathrm{d}V. \qquad (3.38)
\end{aligned}$$

where

$$\left(\vec{\Omega}\cdot\vec{\nabla}\right) = \Omega_x\frac{\partial}{\partial x} + \Omega_y\frac{\partial}{\partial y} + \Omega_z\frac{\partial}{\partial z}.$$

Then, we insert this result into Eq. (3.37), divide the both sides of Eq. (3.37) by ΔV and take the limit $\Delta V \to 0$:

$$\left(\vec{\Omega}\cdot\vec{\nabla}\right)\Phi\left(\vec{r},\vec{\Omega},E\right) = S\left(\vec{r},\vec{\Omega},E\right). \qquad (3.39)$$

Now, in three dimensions, the streaming operator has the form $(\vec{\Omega}\cdot\vec{\nabla})$. It was implied in the above derivation that particles did not interact with matter. If they did, then particle energies and directions would change. Some particles would also be absorbed. None of these processes has been accounted for. Then, we can call Eq. (3.39) the balance equation for a free flow of particles. Next, we will derive a more general balance equation that accounts for particle interactions.

A Simple Form of the Boltzmann Equation

Let us consider the same volume ΔV as in the previous section, but now we will count only particles with a certain momentum $\vec{p}$. Then, compared with the free flow equation we need to add two processes: loss and gain of particles with momentum $\vec{p}$ caused by scattering within ΔV:

{number of particles with momentum $\vec{p}$ entering ΔV}
+ {particles with momentum $\vec{p}$ produced by a source within ΔV}
+ {particles with momentum $\vec{p}$ produced within ΔV by scattering}
= {particles with momentum $\vec{p}$ leaving ΔV}
+ {particles with momentum $\vec{p}$ lost due to scattering within ΔV}.

Particles are lost due to scattering because the momentum changes. If a particle within ΔV had momentum $\vec{p}$ and underwent scattering, $\vec{p} \rightarrow \vec{p}\,'$, then we must subtract the particle, because now it has a "wrong" momentum $\vec{p}\,'$. If instead of scattering the particle is absorbed, this of course will also result in particle loss. To find the total number of particles lost in these two processes, we need to integrate the collision density $\sigma\Phi$ over ΔV. Because σ is the total cross section, it includes both scattering and absorption and therefore does account for both processes. Scattering can also add particles. This happens when a particle within ΔV, with a "wrong" momentum $\vec{p}\,'$, undergoes scattering $\vec{p}\,' \rightarrow \vec{p}$. The collision density for this process is $\sigma_s(\vec{p}\,' \rightarrow \vec{p}\,)\Phi(\vec{p}\,')$. We, again, need to integrate it over ΔV, but also over $\vec{p}\,'$. The latter integral is needed because it co-adds probabilities of $\vec{p}\,' \rightarrow \vec{p}$ scattering for all possible $\vec{p}\,'$.

Adding all the parts together, dividing both sides of the equation by ΔV and taking limit $\Delta V \rightarrow 0$, we arrive at the Boltzmann equation:

$$\left(\vec{\Omega}\cdot\vec{\nabla}\right)\Phi\left(\vec{r},\vec{\Omega},E\right)+\sigma\left(\vec{r},E\right)\Phi\left(\vec{r},\vec{\Omega},E\right) = S\left(\vec{r},\vec{\Omega},E\right)+\int \mathrm{d}E'\int \mathrm{d}\vec{\Omega}\,'\sigma_s\left(\vec{r},\vec{\Omega}\,',E'\rightarrow\vec{\Omega},E\right)\Phi\left(\vec{r},\vec{\Omega}\,',E'\right). \tag{3.40}$$

The first term in the left-hand side of the equation accounts for particle travel, because we derived it by counting particles entering and leaving volume ΔV through its surface. For this reason, $(\vec{\Omega}\cdot\vec{\nabla})$ is called the streaming operator. The next term, $\sigma\Phi$, is the collision density. It is referred to as the removal operator. It accounts for the loss of particles with momentum corresponding to $(\vec{\Omega},E)$ in all types of interaction: scattering, absorption, and any interactions that result in momentum change. In the right-hand side, the source function S characterizes the spatial, angular, and energy distributions of the radiation source. Finally, the last term is called the collision integral. It has the same sign as the source function and is also called the scattering source.

To summarize, the Boltzmann equation is a balance equation. It expresses conservation of the number of particles in phase space. The equation is linear, integro-differential. The linearity is an important property that results from the underlying assumptions that particles do not interact one with another and that material properties, such as the cross sections, do not change when fluence changes. The equation can be conveniently written in an operator form:

$$\hat{L}\Phi = S, \tag{3.41}$$

where

$$\hat{L} = \left(\vec{\Omega}\cdot\vec{\nabla}\right)+\sigma-\hat{K}_s. \tag{3.42}$$

and

$$\hat{K}_s\Phi = \int \mathrm{d}E' \int \mathrm{d}\vec{\Omega}\,' \sigma_s\left(\vec{r}, \vec{\Omega}\,', E' \to \vec{\Omega}, E\right) \Phi\left(\vec{r}, \vec{\Omega}\,', E'\right). \tag{3.43}$$

3.3 Adjoint Transport Equation

The adjoint equation is an equation for the adjoint function $\Phi^+(x)$. In this section we will derive it using a rather formal technique that relies on the rules for adjoint operators summarized in the Appendix B. We start with Eq. (3.34):

$$J = (\Phi, D) = \left(S, \Phi^+\right), \tag{3.44}$$

and then transform its last part:

$$\left(S, \Phi^+\right) = \left(\hat{L}\Phi, \Phi^+\right) = \left(\Phi, \hat{L}^+\Phi^+\right). \tag{3.45}$$

In this transformation, we first used the Boltzmann equation ($S = \hat{L}\Phi$) and then the definition of an adjoint operator. Comparing the last part of Eq. (3.45) with the first half of Eq. (3.44), we can see that

$$\hat{L}^+\Phi^+ = D. \tag{3.46}$$

At this point, we do not know the explicit form of the operator $\hat{L}^+$. What we have determined so far is that it is adjoint to operator $\hat{L}$. Fortunately, we have just derived $\hat{L}$, Eq. (3.42), and finding an adjoint operator in this case is straightforward (see Appendix B), so we can go directly to the result:

$$\hat{L}^+ = -\left(\vec{\Omega}\cdot\vec{\nabla}\right) + \sigma - \hat{K}_s^+, \tag{3.47}$$

where $\hat{K}_s^+$ is identical to the integral defining $\hat{K}_s$, Eq. (3.43), except that it has a transposed kernel, $\sigma_s(\vec{r}, \vec{\Omega}, E \to \vec{\Omega}\,', E')$. Finally, for completeness we present the adjoint equation in a more explicit form:

$$\begin{aligned} &-\left(\vec{\Omega}\cdot\vec{\nabla}\right)\Phi^+\left(\vec{r}, \vec{\Omega}, E\right) + \sigma\left(\vec{r}, E\right)\Phi^+\left(\vec{r}, \vec{\Omega}, E\right) \\ &\quad = D\left(\vec{r}, \vec{\Omega}, E\right) + \int \mathrm{d}E' \int \mathrm{d}\vec{\Omega}\,' \sigma_s\left(\vec{r}, \vec{\Omega}, E \to \vec{\Omega}\,', E'\right)\Phi^+\left(\vec{r}, \vec{\Omega}\,', E'\right). \end{aligned} \tag{3.48}$$

3.4 Overview of the Formalism

In the most general terms, the problem can be formulated as follows: for a given source $S(x)$ and detector $D(x)$, calculate the detector reading, J. The solution depends on the geometry and material properties of objects surrounding the source and detector. As a practical matter, only a finite volume V around the source and detector can be included in calculations. We will call this volume the computational domain. The surface Γ defining that volume should be drawn so as to minimize the computational domain for faster calculations. On the other hand, drawing Γ too close to the source or detector may introduce systematic errors, typically underestimation of J through the elimination of particles scattered back towards the detector by objects outside Γ. In general, Γ should be drawn so that a particle that has left V has a negligible probability of reaching the detector.

Once the boundary Γ is defined, the boundary conditions can be formulated. For a convex surface Γ, the standard boundary condition for fluence is that the fluence incident on Γ from outside is zero. This means that there are no sources outside Γ and that particles that have left V do not scatter back and reenter the volume. For the adjoint function, the boundary condition is that the adjoint function is zero for particles leaving V. In other words, a particle that has left V will never reach the detector and contribute to its reading J. These boundary conditions are sometimes referred to as vacuum boundary conditions. To solve this general problem, we can use one of the two methods:

Boltzmann Equation

$$\hat{L}\Phi = S;\ \ J = (\Phi, D)\,;\ \ \Phi\left(\vec{r}, \vec{\Omega}, E\right) = 0,\ \text{ if } \vec{r} \in \Gamma \ \text{ and } \ \left(\vec{\Omega} \cdot \vec{n}_{\Gamma}\right) < 0. \tag{3.49}$$

Adjoint Transport Equation

$$\hat{L}^{+}\Phi = D; J = \left(S, \Phi^{+}\right); \Phi^{+}\left(\vec{r}, \vec{\Omega}, E\right) = 0, \text{ if} \vec{r} \in \Gamma \text{and } \left(\vec{\Omega} \cdot \vec{n}_{\Gamma}\right) > 0. \tag{3.50}$$

Here, $\vec{n}_{\Gamma}$ is a unit vector, normal to Γ and pointing outwards, Fig. 3.6. It can be seen in Fig. 3.6 that the condition $(\vec{\Omega} \cdot \vec{n}_{\Gamma}) < 0$ identifies particles entering the volume and that $(\vec{\Omega} \cdot \vec{n}_{\Gamma}) > 0$ identifies particles leaving it.

There is an apparent similarity between this general formalism and that of quantum mechanics. Let us take the second formula in Eq. (3.50) and replace S with $\hat{L}\Phi$. This brings us to an expression for the observable J

$$J = \left(\Phi^+, S\right) = \left(\Phi^+, \hat{L}\Phi\right) \tag{3.51}$$

that is very similar to the expression for the mean value of an observable O in quantum mechanics:

$$\langle O \rangle = \langle \psi | \hat{O} | \psi \rangle = \int \psi^* (x)\, \hat{O}\psi (x)\, \mathrm{d}x. \tag{3.52}$$

Here, ψ and ψ^* are the wave function and its complex conjugate, $\hat{O}$ is the operator corresponding to observable O.

The two methods given by Eqs. (3.49) and (3.50) can be combined into a unified formalism. We start by splitting the integral over $\vec{r}$ in the second formula of Eq. (3.50) into two integrals, one over an arbitrary volume V and the other over its complement $\bar{V}$:

$$J = \left(S, \Phi^+\right)_V + \left(S, \Phi^+\right)_{\bar{V}}. \tag{3.53}$$

Then, we will leave the first integral as is, and transform the second integral into an expression similar to (Φ, D). That would be elementary, if we integrated over an infinite volume. In that case, we would have $(S, \Phi^+) = (\Phi, D)$. With a finite integration volume, however, the streaming operator adds an integral over the volume surface. In the integral over $\bar{V}$ we replace S with $\hat{L}\Phi$:

$$\left(S, \Phi^+\right)_{\bar{V}} = \left(\hat{L}\Phi, \Phi^+\right)_{\bar{V}} = \left((\vec{\Omega} \cdot \nabla)\Phi + \sigma\Phi - \hat{K}_s\Phi, \Phi^+\right)_{\bar{V}}. \tag{3.54}$$

Integration over $\vec{r}$ does not affect the removal and scattering operators and for this part we can use the definition of an adjoint operator:

$$\left(\sigma\Phi - \hat{K}_s\Phi, \Phi^+\right)_{\bar{V}} = \left(\Phi, \sigma\Phi^+ - \hat{K}_s^+\Phi^+\right)_{\bar{V}}. \tag{3.55}$$

As for the streaming operator, first we use a differentiation identity:

$$\left((\vec{\Omega} \cdot \nabla)\Phi, \Phi^+\right)_{\bar{V}} = \left((\vec{\Omega} \cdot \nabla)\Phi\Phi^+, 1\right)_{\bar{V}} - \left(\Phi, (\vec{\Omega} \cdot \nabla)\Phi^+\right)_{\bar{V}}. \tag{3.56}$$

Then we add the last term in the above equation to the right-hand side of Eq. (3.55):

$$-\left(\Phi, (\vec{\Omega} \cdot \nabla)\Phi^+\right)_{\bar{V}} + \left(\Phi, \sigma\Phi^+ - \hat{K}_s^+\Phi^+\right)_{\bar{V}} = \left(\Phi, \hat{L}^+\Phi^+\right)_{\bar{V}} = (\Phi, D)_{\bar{V}}, \tag{3.57}$$

where we used Eq. (3.47) for L^+ and then the adjoint equation, Eq. (3.46). In the right-hand side of Eq. (3.56), we transform the remaining volume integral into a surface integral using the divergence theorem:

$$\left((\vec{\Omega}\cdot\nabla)\Phi\Phi^{+},1\right)_{\bar{V}}=\int \mathrm{d}\vec{\Omega}\int \mathrm{d}E\oint_{\Gamma(\bar{V})}\mathrm{d}\Gamma\ \left(\vec{\Omega}\cdot\vec{n}_{\Gamma}\right)\Phi\left(\vec{r},\vec{\Omega},E\right)\Phi^{+}\left(\vec{r},\vec{\Omega},E\right), \tag{3.58}$$

where $\Gamma(\bar{V})$ is the surface of volume $\bar{V}$ and $\vec{n}_{\Gamma}$ is the outward pointing normal to $\Gamma(\bar{V})$. The direction of the normal $\vec{n}_{\Gamma}$ needs clarification. If, for example, V is a sphere, then $\bar{V}$ is everything outside the sphere. The two volumes share the same bounding surface, that is, $\Gamma(V)=\Gamma(\bar{V})$. However, for any given point on this shared surface the outward directions for V and $\bar{V}$ are opposite, that is, $\vec{n}_{\Gamma(V)}=-\vec{n}_{\Gamma(\bar{V})}$.

Finally, after combining Eqs. (3.53), (3.57), and (3.58) we arrive at a representation for detector reading J that unifies the two methods defined by Eqs. (3.49) and (3.50):

$$\begin{aligned}J=&\left(S,\Phi^{+}\right)_{V}+(\Phi,D)_{\bar{V}}\\&+\int \mathrm{d}\vec{\Omega}\int \mathrm{d}E\oint_{\Gamma(\bar{V})}\mathrm{d}\Gamma\ \left(\vec{\Omega}\cdot\vec{n}_{\Gamma}\right)\Phi\left(\vec{r},\vec{\Omega},E\right)\Phi^{+}\left(\vec{r},\vec{\Omega},E\right).\end{aligned} \tag{3.59}$$

The above equation transforms into the adjoint representation, $J=(S,\Phi^{+})$, when volume V is infinite, and the boundary $\Gamma(\bar{V})$ of the complementary volume is at infinity, where both Φ and Φ^{+} are zero. Alternatively, if $V=0$, then Eq. (3.59) becomes $J=(\Phi,D)$. In this case the surface integral is zero, because the surface $\Gamma(\bar{V})$ vanishes.

3.5 The Lagrangian Form of the Boltzmann Equation

In Sect. 3.2, when we derived the Boltzmann equation, we noted strong parallels with fluid mechanics. In this section, we rely on terminology adapted from fluid mechanics to introduce another form of the Boltzmann equation, that we will call the Lagrangian form. This didactic approach and the terminology that we use, although unconventional, offer an accurate description of the meaning of the equation, and a clearer perspective. Unconventional does not mean new. For example, Wienke (1982) in a paper on an electron transport code derived a transport equation that he described as Lagrangian.

We derived the Boltzmann equation in a static, laboratory frame. From this equation, for an arbitrary point $\vec{r}$ in space, or a point in phase space, $(\vec{r},\vec{\Omega},E)$, we can, in principle, find the fluence. This approach to studying a field, in this case the fluence field $\Phi(x)$, corresponds to the Eulerian description of fluid flow. An alternative description of fluid flow is called Lagrangian (Granger 1995). In the Lagrangian formalism, the reference frame is associated with or, in other words, "attached" to a fluid parcel. In radiation transport, it is more convenient to fix the frame to an individual particle than to fix the frame to a parcel. To put it simply, in the Eulerian approach, we choose an arbitrary point, and then calculate the fluence at that point. In the Lagrangian approach, we follow the particle's path and study how

the fluence changes along this path. In Monte Carlo algorithms, we also usually follow the particle's path as we generate, step-by-step, the particle trajectory. This makes the Lagrangian description particularly useful in the context of Monte Carlo simulation of radiation transport.

To derive the Lagrangian form of the Boltzmann equation we consider a particle that after traveling a distance t has arrived at a point $\vec{r} = \vec{r}(t)$. From this point, the particle travels a small distance Δt in direction $\vec{\Omega}$. If no interactions, scattering or absorption, occur as the particle travels the distance Δt, then it arrives at point $\vec{r} + \vec{\Omega}\Delta t$. This expression implies a straight-line trajectory and therefore is valid only in the limit $\Delta t \to 0$. We will omit all the sources, including the collisional source (i.e., the collision integral), and the particle energy, E. This means that we will derive only the left-hand side of the Boltzmann equation. Then, the derivation is simple:

$$\begin{aligned}
&\Phi\left(\vec{r} + \vec{\Omega}\Delta t\,; t + \Delta t\right) \approx \Phi\left(\vec{r}\,; t\right) + \Delta t\left(\vec{\Omega}\cdot\vec{\nabla}\right)\Phi\left(\vec{r}\,; t\right) + \Delta t\frac{\partial}{\partial t}\Phi\left(\vec{r}\,; t\right);\\
&\Phi\left(\vec{r} + \vec{\Omega}\Delta t\,; t + \Delta t\right) \approx \Phi\left(\vec{r}\,; t\right)(1 - \sigma\Delta t);\quad (\Delta t \to 0);\\
&\Rightarrow \frac{\partial}{\partial t}\Phi\left(\vec{r}, t\right) + \left(\vec{\Omega}\cdot\vec{\nabla}\right)\Phi\left(\vec{r}, t\right) + \sigma\Phi\left(\vec{r}, t\right) = 0.
\end{aligned} \tag{3.60}$$

The first line in Eq. (3.60) is the Taylor series expansion of the fluence. It is valid whether or not the particle actually arrives at point $\vec{r} + \vec{\Omega}\Delta t$. The second line accounts for the possibility of particle loss, which occurs with the probability $\sigma\Delta t$. In the last line, the first two lines are combined and the limit $\Delta t \to 0$ is taken. Comparing this result with the left-hand side of the Boltzmann equation, Eq. (3.40), we can see that in the Lagrangian form, the equation has an extra term in the left-hand side, the derivative $\partial\Phi/\partial t$. To account for the collisional and external sources, we would have to add them, in an appropriate form, to the right-hand side of the second equation in Eq. (3.60). This would produce exactly the same expression in the right-hand side of the Lagrangian form as we had in the original form of the Boltzmann equation, Eq. (3.40). A complete derivation of the Lagrangian transport equation is given in Wienke (1974, 1982). An essentially the same equation, although not described as Lagrangian, was derived by a different method by Kolchuzhkin and Uchaikin (1978).

Finally, in fluid mechanics each fluid parcel is usually labeled so as to distinguish it from other parcels. We will label the particle by its point of origin in the phase space, $(\vec{r}_0, \vec{\Omega}_0, E_0)$. Also, because we are interested in a solution only for $t > 0$, we can leave out the source function S. Then, the Boltzmann equation in the Lagrangian form is:

$$\frac{\partial}{\partial t}\Phi\left(\vec{r},\vec{\Omega},E,t|\vec{r}_0,\vec{\Omega}_0,E_0\right)+\left(\vec{\Omega}\cdot\vec{\nabla}\right)\Phi\left(\vec{r},\vec{\Omega},E,t|\vec{r}_0,\vec{\Omega}_0,E_0\right)$$
$$+\sigma\left(\vec{r},E\right)\Phi\left(\vec{r},\vec{\Omega},E,t|\vec{r}_0,\vec{\Omega}_0,E_0\right)$$
$$=\int \mathrm{d}E'\int \mathrm{d}\vec{\Omega}'\sigma_s\left(\vec{r},\vec{\Omega}',E'\rightarrow\vec{\Omega},E\right)\Phi\left(\vec{r},\vec{\Omega}',E',t|\vec{r}_0,\vec{\Omega}_0,E_0\right). \tag{3.61}$$

Equation (3.61) is the basis of multiple scattering models used for generating trajectories of charged particles in condensed history algorithms. This topic is discussed in Chap. 5.

3.6 The Boltzmann Equation for Multiplying Systems

For simplicity, we consider a system that has only two types of interactions. The first type is scattering, during which the particle momentum changes from $\vec{p}\,'$ to $\vec{p}$, and no secondary particles are produced. The cross section of this process is $\sigma_1(\vec{p}\,'\rightarrow\vec{p})$. The second type is particle multiplication, during which the momentum of the incident particle changes, and a secondary particle of the same type as the incident particle is produced. The cross section of this process is $\sigma_2(\vec{p}\,'\rightarrow\vec{p}_1,\vec{p}_2)$, where $\vec{p}_1$ and $\vec{p}_2$ are the momenta of the two particles emerging from the collision. In the case of electron transport, an example of the first process is excitation, and an example of the second process is ionization. The Boltzmann equation for this simple multiplying system is the same as Eq. (3.40), except for an additional collision integral that accounts for the second interaction type:

$$\left(\vec{\Omega}\cdot\nabla\right)\Phi\left(\vec{p}\,\right)+\sigma\Phi\left(\vec{p}\,\right)-\int \mathrm{d}\vec{p}\,'\sigma_1\left(\vec{p}\,'\rightarrow\vec{p}\,\right)\Phi\left(\vec{p}\,'\right)$$
$$-\int \mathrm{d}\vec{p}\,'\,\Phi\left(\vec{p}\,'\right)\int \mathrm{d}\vec{p}\,''\left[\sigma_2\left(\vec{p}\,'\rightarrow\vec{p},\vec{p}\,''\right)+\sigma_2\left(\vec{p}\,'\rightarrow\vec{p}\,'',\vec{p}\,\right)\right]=S\left(\vec{p}\,\right). \tag{3.62}$$

Here, for brevity, we omitted the variable $\vec{r}$, and switched from the usual variables $(\vec{\Omega},E)$ to momentum $\vec{p}$. The meaning of the additional collision integral is obvious. The first term in the square brackets accounts for collisions that result in the first of the two particles emerging with momentum $\vec{p}$. The second term accounts for those collisions from which the second particle emerges with momentum $\vec{p}$. This equation can be simplified by integrating over $\vec{p}\,''$

$$\int \mathrm{d}\vec{p}\,''\left[\sigma_2\left(\vec{p}\,'\rightarrow\vec{p},\vec{p}\,''\right)+\sigma_2\left(\vec{p}\,'\rightarrow\vec{p}\,'',\vec{p}\,\right)\right]=\tilde{\sigma}_2\left(\vec{p}\,'\rightarrow\vec{p}\,\right). \tag{3.63}$$

3.7 Adjoint Transport Equation for Multiplying Systems

The adjoint transport equation for nonmultiplying systems was derived in Sect. 3.3. To extend it to the simplified multiplying system that we introduced in the preceding section, we need to find the operator adjoint to the scattering operator in the second collision integral in Eq. (3.62). To do so, we consider the following integral:

$$\begin{aligned} I &= \left(\Phi^+, \hat{K}_2 \Phi\right) \\ &\equiv \int \mathrm{d}\vec{p}\ \Phi^+\left(\vec{p}\,\right) \int \mathrm{d}\vec{p}\,'\, \Phi\left(\vec{p}\,'\right) \int \mathrm{d}\vec{p}\,''\left[\sigma_2\left(\vec{p}\,' \to \vec{p}, \vec{p}\,''\right) + \sigma_2\left(\vec{p}\,' \to \vec{p}\,'', \vec{p}\,\right)\right]. \end{aligned} \tag{3.64}$$

Next, we rename the integration variables $\vec{p} \to \vec{p}\,'$, $\vec{p}\,' \to \vec{p}$, and change the order of integration

$$I = \int \mathrm{d}\vec{p}\ \Phi\left(\vec{p}\,\right) \int \mathrm{d}\vec{p}\,'\, \Phi^+\left(\vec{p}\,'\right) \int \mathrm{d}\vec{p}\,''\left[\sigma_2\left(\vec{p} \to \vec{p}\,', \vec{p}\,''\right) + \sigma_2\left(\vec{p} \to \vec{p}\,'', \vec{p}\,'\right)\right]. \tag{3.65}$$

Finally, we move Φ^+ within the square brackets, and in the integral involving the last term in the square brackets, rename variables: $\vec{p}\,' \to \vec{p}\,''$ and $\vec{p}\,'' \to \vec{p}\,'$:

$$\begin{aligned} I &= \int \mathrm{d}\vec{p}\ \Phi\left(\vec{p}\,\right) \int \mathrm{d}\vec{p}\,' \int \mathrm{d}\vec{p}\,''\sigma_2\left(\vec{p} \to \vec{p}\,', \vec{p}\,''\right)\left[\Phi^+\left(\vec{p}\,'\right) + \Phi^+\left(\vec{p}\,''\right)\right] \\ &= \left(\Phi, \hat{K}_2^+ \Phi^+\right). \end{aligned} \tag{3.66}$$

The last two integrals in Eq. (3.66) represent the adjoint collision integral for the process described by the cross section σ_2. The adjoint collision integral for the cross section σ_1 is the same as the collision integral $\hat{K}_s^+$ derived in Sect. 3.3. Hence, the adjoint transport equation for the multiplying system is

$$\begin{aligned} &-\left(\vec{\Omega}\cdot\nabla\right)\Phi^+\left(\vec{p}\,\right) + \sigma\Phi^+\left(\vec{p}\,\right) - \int \mathrm{d}\vec{p}\,'\sigma_1\left(\vec{p} \to \vec{p}\,'\right)\Phi^+\left(\vec{p}\,'\right) \\ &\quad - \int \mathrm{d}\vec{p}\,' \int \mathrm{d}\vec{p}\,''\sigma_2\left(\vec{p} \to \vec{p}\,', \vec{p}\,''\right)\left[\Phi^+\left(\vec{p}\,'\right) + \Phi^+\left(\vec{p}\,''\right)\right] = D\left(\vec{p}\,\right). \end{aligned} \tag{3.67}$$

The left-hand side of this equation defines the adjoint operator $\hat{L}^+$.

3.8 The Boltzmann Equation in the Presence of an External Magnetic Field

The method we used in Sect. 3.2 to derive the Boltzmann equation was based on a consideration of the balance of particles in a small volume in phase space. In the presence of an external magnetic field, this method remains valid, but becomes much less intuitive. In this section we present a more formal method based on fundamental equations of nonequilibrium statistical mechanics.

Let us consider a system composed of N particles, $N \gg 1$. The state of the system is characterized by the phase coordinate of the system $x = \{\vec{r}_1, \ldots, \vec{r}_N; \vec{p}_1, \ldots, \vec{p}_N\}$, where $(\vec{r}_i, \vec{p}_i)$ are the phase coordinates of individual particles. The state x of the system is a random quantity. Its probability density function, $F_N(x, t)$, is called the phase space distribution function. In disequilibrium, it is a function of time t. It satisfies the Liouville equation (Balescu 1975). For our system it has the following form:

$$\begin{aligned}\frac{\mathrm{d}F_N}{\mathrm{d}t} &= \frac{\partial F_N}{\partial t} + \sum_{i=1}^{N}\left(\frac{\partial F_N}{\partial x_i}\frac{\mathrm{d}x_i}{\mathrm{d}t} + \frac{\partial F_N}{\partial y_i}\frac{\mathrm{d}y_i}{\mathrm{d}t} + \frac{\partial F_N}{\partial z_i}\frac{\mathrm{d}z_i}{\mathrm{d}t}\right) \\ &+ \sum_{i=1}^{N}\left(\frac{\partial F_N}{\partial p_{ix}}\frac{\mathrm{d}p_{ix}}{\mathrm{d}t} + \frac{\partial F_N}{\partial p_{iy}}\frac{\mathrm{d}p_{iy}}{\mathrm{d}t} + \frac{\partial F_N}{\partial p_{iz}}\frac{\mathrm{d}p_{iz}}{\mathrm{d}t}\right) \\ &= \frac{\partial F_N}{\partial t} + \sum_{i=1}^{N}\left[\left(\vec{v}_i \cdot \vec{\nabla}_r^{(i)}\right) F_N + \left(\vec{f}_i \cdot \vec{\nabla}_p^{(i)}\right) F_N\right] = 0,\end{aligned} \tag{3.68}$$

where $\vec{f}_i$ is the force exerted on the i-th particle by external fields and by other particles, and $\vec{v}_i$ is the particle velocity. The equation means simply that the distribution function is constant along the phase trajectories (Liouville's theorem). Next, we derive from Eq. (3.68) the first equation of the Bogoliubov–Born–Green–Kirkwood–Yvon (BBGKY) hierarchy (Balescu 1975). We introduce two reduced distribution functions:

$$F_1\left(\vec{r}_1, \vec{p}_1\right) \equiv \int \mathrm{d}\vec{r}_2 \ldots \int \mathrm{d}\vec{r}_N \int \mathrm{d}\vec{p}_2 \ldots \int \mathrm{d}\vec{p}_N \, F_N\left(\vec{r}_1, \ldots \vec{r}_N; \vec{p}_1, \ldots \vec{p}_N\right), \tag{3.69}$$

and

$$F_2\left(\vec{r}_1, \vec{r}_2; \vec{p}_1, \vec{p}_2\right) \equiv \int \mathrm{d}\vec{r}_3 \ldots \int \mathrm{d}\vec{r}_N \int \mathrm{d}\vec{p}_3 \ldots \int \mathrm{d}\vec{p}_N \, F_N\left(\vec{r}_1, \ldots \vec{r}_N; \vec{p}_1, \ldots \vec{p}_N\right). \tag{3.70}$$

Then, we integrate Eq. (3.68) over the phase coordinates of particles $2, 3, \ldots N$. We will perform the integration one term at a time:

$$\int \mathrm{d}\vec{r}_2 \ldots \int \mathrm{d}\vec{r}_N \int \mathrm{d}\vec{p}_2 \ldots \int \mathrm{d}\vec{p}_N \, \frac{\partial F_N}{\partial t} = \frac{\partial F_1}{\partial t}. \tag{3.71}$$

$$\int \mathrm{d}\vec{r}_2 \ldots \int \mathrm{d}\vec{r}_N \int \mathrm{d}\vec{p}_2 \ldots \int \mathrm{d}\vec{p}_N \, \left(\vec{v}_i \cdot \vec{\nabla}_r^{(i)}\right) F_N = \begin{cases} \left(\vec{v}_1 \cdot \vec{\nabla}_r^{(1)}\right) F_1; & \text{if } i = 1; \\ 0; & \text{if } i > 1. \end{cases} \tag{3.72}$$

For $i > 1$, the integral in Eq. (3.72) is equal to zero, because by applying the divergence theorem it can be transformed into a surface integral. We place this surface at infinity, where the distribution function is zero.

In the next term, we separate force $\vec{f}_i$ into two forces:

$$\vec{f}_i = \vec{f}_{i,\mathrm{ex}}\left(\vec{r}_i, \vec{p}_i\right) + \sum_{\substack{j=1 \\ (j \neq i)}}^{N} \vec{f}_{ij}\left(|\vec{r}_i - \vec{r}_j|\right), \tag{3.73}$$

where $\vec{f}_{i,\mathrm{ex}}$ is the force of an external field. It depends on the location $\vec{r}_i$ of the particle and its momentum (or velocity) in the case of a magnetic field. In the second term, $\vec{f}_{ij}$ is the force exerted on particle i by particle j. This force depends on the distance $|\vec{r}_i - \vec{r}_j|$ between the two particles, assuming here that the force is that of Coulomb interaction between charged particles. The integral of the term with the external force is similar to the integral in Eq. (3.72), except that the divergence theorem is now applied in momentum space:

$$\begin{aligned} &\int \mathrm{d}\vec{r}_2 \ldots \int \mathrm{d}\vec{r}_N \int \mathrm{d}\vec{p}_2 \ldots \int \mathrm{d}\vec{p}_N \, \left(\vec{f}_{i,\mathrm{ex}} \cdot \vec{\nabla}_p^{(i)}\right) F_N \\ &= \begin{cases} \left(\vec{f}_{1,\mathrm{ex}} \cdot \vec{\nabla}_p^{(1)}\right) F_1, & \text{if } i = 1; \\ 0; & \text{if } i > 1. \end{cases} \end{aligned} \tag{3.74}$$

Similarly, the integral with force f_{ij} is nonzero only for $i = 1$:

$$\begin{aligned} &\int \mathrm{d}\vec{r}_2 \ldots \int \mathrm{d}\vec{r}_N \int \mathrm{d}\vec{p}_2 \ldots \int \mathrm{d}\vec{p}_N \, \left(\vec{f}_{ij} \cdot \vec{\nabla}_p^{(i)}\right) F_N \\ &= \begin{cases} \int \mathrm{d}\vec{r}_j \int \mathrm{d}\vec{p}_j \left(\vec{f}_{1j} \cdot \vec{\nabla}_p^{(1)}\right) F_2\left(\vec{r}_1, \vec{r}_j; \vec{p}_1, \vec{p}_j\right), & \text{if } i = 1; \\ 0, & \text{if } i > 1. \end{cases} \end{aligned} \tag{3.75}$$

For all $j = 1, 2 \ldots i-1, i+1, \ldots N$ the integral is the same. Hence, integrating the sum over j in Eq. (3.73) will produce $N - 1$ integrals of this form. After combining Eqs. (3.71)–(3.75), we arrive at the following form of the first equation of the BBGKY hierarchy:

$$\frac{\partial F_1\left(\vec{r}_1,\vec{p}_1\right)}{\partial t}+\left(\vec{f}_{1,\mathrm{ex}}\left(\vec{r}_1,\vec{p}_1\right)\cdot\vec{\nabla}_p^{(1)}\right)F_1\left(\vec{r}_1,\vec{p}_1\right)+\left(\vec{v}_1\cdot\vec{\nabla}_r^{(1)}\right)F_1\left(\vec{r}_1,\vec{p}_1\right)$$
$$+(N-1)\int \mathrm{d}\vec{r}_2\int \mathrm{d}\vec{p}_2\left(\vec{f}_{12}\left(|\vec{r}_1-\vec{r}_2|\right)\cdot\vec{\nabla}_p^{(1)}\right)F_2\left(\vec{r}_1,\vec{r}_2;\vec{p}_1,\vec{p}_2\right)=0. \tag{3.76}$$

The integral term in this equation accounts for interactions between particles. The key approximation for simplifying this integral is the molecular chaos hypothesis, which assumes statistical independence of particles. This approximation leads to the well-known Boltzmann kinetic equation for dilute gases. This equation is nonlinear, because the particles, in this case gas molecules, interact one with another. The integral can be written in terms of the interaction cross section. When this approach is applied to radiation transport, a further simplification can be made that the particles do not interact one with another. They interact with matter, and do so in such a way that changes in the material properties caused by the interactions, such as ionizations of individual atoms, have only a negligible effect on radiation transport. This makes the Boltzmann equation linear. These derivations are lengthy, and we will omit them. Our focus is on including an external magnetic field, which is represented in Eq. (3.76) by the force $f_{1,\mathrm{ex}}$. As for the integral term, after the above simplifications are made, it will produce the removal operator and the same collision integral as in Eq. (3.40).

To simplify Eq. (3.76) and bring it into a form similar to Eq. (3.40), we will (a) consider a steady state form of the equation, and set the derivative $\partial F_1/\partial t$ to zero; (b) omit the subscript "1"; (c) assume a uniform magnetic field, so that $f_{1,\mathrm{ex}}(\vec{r}_1,\vec{p}_1)=f_{1,\mathrm{ex}}(\vec{p}_1)$; (d) write Eq. (3.19) (definition 2 of fluence) as $\Phi(\vec{r}_1,\vec{p}_1)=nF_1(\vec{r}_1,\vec{p}_1)v_1\Delta t$, and then use it to replace F_1 with the fluence; the average number density n and Δt will ultimately cancel out; (e) use $\vec{v}=v\vec{\Omega}$; (f) replace the integral term with the removal operator and the collision integral; and (g) add the source function S. Then, Eq. (3.76) takes the following form:

$$\left(\vec{f}_{\mathrm{ex}}\left(\vec{p}\,\right)\cdot\vec{\nabla}_p\right)\left[\frac{1}{v}\Phi\left(\vec{r},\vec{p}\,\right)\right]+\left(\vec{\Omega}\cdot\vec{\nabla}_r\right)\Phi\left(\vec{r},\vec{p}\,\right)+\sigma\left(\vec{r},E\right)\Phi\left(\vec{r},\vec{p}\,\right)$$
$$=S\left(\vec{r},\vec{p}\,\right)+\int \mathrm{d}\vec{p}\,'\sigma_s\left(\vec{r},\vec{p}\,'\rightarrow\vec{p}\,\right)\Phi\left(\vec{r},\vec{p}\,'\right). \tag{3.77}$$

Thus, when an external force is present, a new term must be added to the Boltzmann equation. The term has a form similar to the streaming operator. It can be described as a streaming operator in the momentum space. The force of a magnetic field $\vec{B}$ on a particle of charge q and velocity $\vec{v}$ is given by the Lorentz formula (Jackson 1999)

$$\vec{f}_{\mathrm{ex}}\left(\vec{p}\,\right)=q\,\vec{v}\times\vec{B}. \tag{3.78}$$

The force of an electric field will not be considered. The electric field $\vec{E}$ is set to zero.

Our approach has been based on the Liouville equation. On the other hand, we previously emphasized the similarity between the Boltzmann equation and the continuity equation. The Liouville equation, however, is not exactly equivalent to the continuity equation. Let us examine the difference between the two equations. In three dimensions the steady state continuity equation, in our notation, is:

$$\vec{\nabla}_r \cdot \left[\vec{v}\, F_1\left(\vec{r}\,\right)\right] = \frac{\partial}{\partial x}\left[F_1\left(\vec{r}\,\right)\frac{\mathrm{d}x}{\mathrm{d}t}\right] + \frac{\partial}{\partial y}\left[F_1\left(\vec{r}\,\right)\frac{\mathrm{d}y}{\mathrm{d}t}\right] + \frac{\partial}{\partial z}\left[F_1\left(\vec{r}\,\right)\frac{\mathrm{d}z}{\mathrm{d}t}\right] = 0. \tag{3.79}$$

Extending this equation to six-dimensional phase space is straightforward:

$$\begin{aligned}\frac{\partial}{\partial x}\left[F_1\left(\vec{r},\vec{p}\,\right)\frac{\mathrm{d}x}{\mathrm{d}t}\right] + \frac{\partial}{\partial y}\left[F_1\left(\vec{r},\vec{p}\,\right)\frac{\mathrm{d}y}{\mathrm{d}t}\right] + \frac{\partial}{\partial z}\left[F_1\left(\vec{r},\vec{p}\,\right)\frac{\mathrm{d}z}{\mathrm{d}t}\right] \\ + \frac{\partial}{\partial p_x}\left[F_1\left(\vec{r},\vec{p}\,\right)\frac{\mathrm{d}p_x}{\mathrm{d}t}\right] + \frac{\partial}{\partial p_y}\left[F_1\left(\vec{r},\vec{p}\,\right)\frac{\mathrm{d}p_y}{\mathrm{d}t}\right] + \frac{\partial}{\partial p_z}\left[F_1\left(\vec{r},\vec{p}\,\right)\frac{\mathrm{d}p_z}{\mathrm{d}t}\right] = 0.\end{aligned} \tag{3.80}$$

Written in terms of the fluence, this result is similar to the left-hand side of Eq. (3.77):

$$\left(\vec{\Omega}\cdot\vec{\nabla}_r\right)\Phi\left(\vec{r},\vec{p}\,\right) + \vec{\nabla}_p \cdot \left(\vec{f}_{\mathrm{ex}}\left(\vec{p}\,\right)\frac{1}{v}\Phi\left(\vec{r},\vec{p}\,\right)\right) = 0, \tag{3.81}$$

except that, Eq. (3.81) has a different form of the term accounting for the external force. To evaluate the difference, we can use a vector identity (Jackson 1999):

$$\vec{\nabla}\cdot\left(\psi\vec{a}\,\right) = \vec{a}\cdot\vec{\nabla}\psi + \psi\vec{\nabla}\cdot\vec{a} \tag{3.82}$$

Then, the difference between the external force terms in Eqs. (3.81) and (3.77) is

$$\vec{\nabla}_p \cdot \left(\vec{f}_{\mathrm{ex}}\left(\vec{p}\,\right)\frac{1}{v}\Phi\left(\vec{r},\vec{p}\,\right)\right) - \left(\vec{f}_{\mathrm{ex}}\left(\vec{p}\,\right)\cdot\vec{\nabla}_p\right)\frac{1}{v}\Phi\left(\vec{r},\vec{p}\,\right) = \frac{1}{v}\Phi\left(\vec{r},\vec{p}\,\right)\left(\vec{\nabla}_p\cdot\vec{f}_{\mathrm{ex}}\left(\vec{p}\,\right)\right). \tag{3.83}$$

Then, we insert the expression for Lorentz force into the right-hand side of Eq. (3.83), and use another vector identity (Jackson 1999):

$$\vec{\nabla}\cdot\left[\vec{a}\times\vec{b}\,\right] = \vec{b}\cdot\left[\vec{\nabla}\times\vec{a}\,\right] - \vec{a}\cdot\left[\vec{\nabla}\times\vec{b}\,\right]. \tag{3.84}$$

This yields:

$$\left(\vec{\nabla}_p\cdot\vec{f}_{\mathrm{ex}}\left(\vec{p}\,\right)\right) = q\left(\vec{\nabla}_p\cdot\left[\vec{v}\times\vec{B}\,\right]\right) = q\left(\vec{B}\cdot\left[\vec{\nabla}_p\times\vec{v}\,\right]\right) - q\left(\vec{v}\cdot\left[\vec{\nabla}_p\times\vec{B}\,\right]\right) = 0. \tag{3.85}$$

The expression $[\vec{\nabla}_p \times \vec{v}\ ]$ is identically zero, and $[\vec{\nabla}_p \times \vec{B}\,]$ is also zero, because the magnetic field does not depend on momentum. This shows the equivalency of

the Liouville equation and the continuity equation when only a magnetic field is present. In more general language, the two equations are equivalent, if the system in question satisfies the Hamiltonian equations of motion (Landau and Lifshitz 1969). A relevant example of a non-Hamiltonian system is such a system of charged particles where the radiative damping force is not negligible (Polyakov 1988).

3.9 Simplified Forms of the Boltzmann Equation

3.9.1 Unscattered Fluence

The unscattered fluence Φ_0 is an important quantity. It is present in the right-hand side of the integral form of the Boltzmann equation, Eq. (3.191). In grid-based algorithms (Chap. 7), the first step is usually calculation of unscattered fluence. To derive an equation for Φ_0, in the Boltzmann equation, Eq. (3.40), we set the scattering cross section σ_s to zero:

$$\left(\vec{\Omega}\cdot\vec{\nabla}\right)\Phi_0\left(\vec{r},\vec{\Omega},E\right)+\sigma\left(\vec{r},E\right)\Phi_0\left(\vec{r},\vec{\Omega},E\right)=S\left(\vec{r},\vec{\Omega},E\right). \tag{3.86}$$

This equation can be solved analytically. For the boundary Γ of the computational domain, we will use the following boundary condition:

$$\Phi_0\left(\vec{r},\vec{\Omega},E\right)=\Phi_\Gamma\left(\vec{r},\vec{\Omega},E\right),\ \text{if}\ \vec{r}\in\Gamma\ \text{and}\ \left(\vec{\Omega}\cdot\vec{n}_\Gamma\right)<0. \tag{3.87}$$

where $\vec{n}_\Gamma$ is an outward-pointing normal to Γ. This is a slightly more general form of the boundary conditions than the vacuum condition we introduced earlier in Eq. (3.49). It means that there is a fluence Φ_Γ of particles incident from outside on Γ, implying that not all radiation sources are located within the computational domain. Setting $\Phi_\Gamma = 0$ in Eq. (3.87) produces the standard vacuum boundary conditions.

The streaming operator can be considered as a directional derivative with respect to direction $\vec{\Omega}$, which can be written in the following form:

$$\left(\vec{\Omega}\cdot\vec{\nabla}\right)\Phi_0\left(\vec{r}-\vec{\Omega}t,\vec{\Omega},E\right)=-\frac{\partial}{\partial t}\Phi_0\left(\vec{r}-\vec{\Omega}t,\vec{\Omega},E\right). \tag{3.88}$$

This equation is written for all points on a straight line defined by the equation $\vec{r}-\vec{\Omega}t$, including point $\vec{r}$ when $t=0$. This transforms Eq. (3.86) into a differential equation in terms of variable t:

$$-\frac{\partial}{\partial t}\Phi_0(t)+\sigma(t)\,\Phi_0(t)=S(t). \tag{3.89}$$

with the boundary condition:

$$\Phi_0\left(t_{\max}\right) = \Phi_\Gamma\left(\vec{r} - \vec{\Omega} t_{\max}, \vec{\Omega}, E\right). \tag{3.90}$$

where $t_{\max}$ is the distance from $\vec{r}$ to surface Γ in the direction opposite to $\vec{\Omega}$. We will solve this differential equation using the method of undetermined coefficients. We will seek a solution in the following form:

$$\Phi_0\left(t\right) = A\left(t\right) \exp\left[\int_0^t \sigma\left(t'\right) \mathrm{d}t'\right]. \tag{3.91}$$

We found this form by solving Eq. (3.89) with its right-hand side, $S(t)$, set to zero. To determine coefficient $A(t)$, we substitute Eq. (3.91) in Eq. (3.89), which leads to:

$$\frac{\partial}{\partial t} A\left(t\right) = -S\left(t\right) \exp\left[-\int_0^t \sigma\left(t'\right) \mathrm{d}t'\right]. \tag{3.92}$$

Next, we integrate this equation from $t_{\max}$ to t:

$$A\left(t\right) = A\left(t_{\max}\right) + \int_t^{t_{\max}} S\left(t''\right) \exp\left[-\int_0^{t''} \sigma\left(t'\right) \mathrm{d}t'\right] \mathrm{d}t''. \tag{3.93}$$

From Eqs. (3.90) and (3.91) we find

$$A\left(t_{\max}\right) = \Phi_\Gamma\left(\vec{r} - \vec{\Omega} t_{\max}, \vec{\Omega}, E\right) \exp\left[-\int_0^{t_{\max}} \sigma\left(t'\right) \mathrm{d}t'\right]. \tag{3.94}$$

Finally, we substitute Eqs. (3.93) and (3.94) in Eqs. (3.91), set $t = 0$, and arrive at the solution:

$$\begin{aligned}\Phi_0\left(\vec{r}, \vec{\Omega}, E\right) = {} & \Phi_\Gamma\left(\vec{r} - \vec{\Omega} t_{\max}, \vec{\Omega}, E\right) \exp\left[-\int_0^{t_{\max}} \sigma\left(\vec{r} - \vec{\Omega} t', E\right) \mathrm{d}t'\right] \\ & + \int_0^{t_{\max}} S\left(\vec{r} - \vec{\Omega} t'', \vec{\Omega}, E\right) \exp\left[-\int_0^{t''} \sigma\left(\vec{r} - \vec{\Omega} t', E\right) \mathrm{d}t'\right] \mathrm{d}t''.\end{aligned} \tag{3.95}$$

Note that the integrals in square brackets are optical distances. For example,

$$\int_0^{t''} \sigma\left(\vec{r} - \vec{\Omega} t', \vec{\Omega}, E\right) \mathrm{d}t' = \tau\left(\vec{r}, \vec{r}', \vec{\Omega}, E\right), \tag{3.96}$$

where $\vec{r}' = \vec{r} - \vec{\Omega} t''$.

Example 1

A point mono-directional source in an infinite homogeneous medium ($t_{\max} = \infty$, $\sigma(t, E) = \sigma(E)$), and $\Phi_\Gamma = 0$. The source function is a product of two δ-functions and an energy distribution function, $S_E(E)$:

$$S\left(\vec{r}, \vec{\Omega}, E\right) = \delta\left(\vec{r} - \vec{r}_0\right) \delta\left(\vec{\Omega} - \vec{\Omega}_0\right) S_E(E) . \tag{3.97}$$

We choose the coordinate system so that the z-axis is parallel to $\vec{\Omega}_0$. Then, $\vec{\Omega}_0 = (0, 0, 1)$, and Eq. (3.95) becomes:

$$\Phi_0\left(\vec{r}, \vec{\Omega}, E\right) = \int_0^\infty \delta(x - x_0)\, \delta(y - y_0)\, \delta\left(z - t'' - z_0\right) \exp\left[-\sigma(E)\, t''\right] \mathrm{d}t'' \times \delta\left(\vec{\Omega} - \vec{\Omega}_0\right) S_E(E) . \tag{3.98}$$

After integrating over t'' we arrive at the final result. For $z \geq z_0$ we have:

$$\Phi_0\left(\vec{r}, \vec{\Omega}, E\right) = \delta(x - x_0)\, \delta(y - y_0) \exp\left[-\sigma(E)(z - z_0)\right] \delta\left(\vec{\Omega} - \vec{\Omega}_0\right) S_E(E) . \tag{3.99}$$

For $z < z_0$ the unscattered fluence is zero. For $z > z_0$ this solution is a narrow monodirectional beam ($x = x_0$, $y = y_0$, $\vec{\Omega} = \vec{\Omega}_0$) attenuated exponentially with increasing distance from the source.

Example 2

A point, isotropically emitting source in an infinite homogeneous medium ($t_{\max} = \infty$, $\sigma(t, E) = \sigma(E)$), and $\Phi_\Gamma = 0$. In this case the source function is:

$$S\left(\vec{r}, \vec{\Omega}, E\right) = \delta\left(\vec{r} - \vec{r}_0\right) \frac{1}{4\pi} S_E(E) . \tag{3.100}$$

We will place the origin of the coordinate system at the location of the source, so that $\vec{r}_0 = (0, 0, 0)$. In this case, to calculate the integral in Eq. (3.95) we will use Eq. (B4) of Appendix B, which transforms a line integral over t into a volume integral over $\vec{r}\,'$, where $\vec{r}\,' = \vec{r} - \vec{\Omega} t''$. We note that the optical distance in the exponent is $\sigma t''$, and we need to express t in terms of the integration variable $\vec{r}\,'$:

$$\vec{\Omega} t'' = \vec{r} - \vec{r}\,' \;\Rightarrow\; t'' = \vec{\Omega}\left(\vec{r} - \vec{r}\,'\right) . \tag{3.101}$$

It follows then that

$$\Phi_0\left(\vec{r},\vec{\Omega},E\right)=\int_0^\infty S\left(\vec{r}-\vec{\Omega}t'',\vec{\Omega},E\right)\exp\left[-\sigma\left(E\right)t''\right]\mathrm{d}t''$$
$$=\int S\left(\vec{r}\,',\vec{\Omega},E\right)\exp\left[-\sigma\left(E\right)\vec{\Omega}\cdot\left(\vec{r}-\vec{r}\,'\right)\right]\delta\left(\vec{\Omega}-\frac{\vec{r}-\vec{r}\,'}{|\vec{r}-\vec{r}\,'|}\right)\frac{\mathrm{d}\vec{r}\,'}{|\vec{r}-\vec{r}\,'|^2}. \tag{3.102}$$

We can now substitute here the source function, Eq. (3.100). Integration over $\vec{r}\,'$ results in replacing $\vec{r}\,'$ with $\vec{r}_0=(0,0,0)$:

$$\Phi_0\left(\vec{r},\vec{\Omega},E\right)=\exp\left[-\sigma\left(E\right)\vec{\Omega}\cdot\vec{r}\,\right]\delta\left(\vec{\Omega}-\frac{\vec{r}}{r}\right)\frac{1}{r^2}\frac{1}{4\pi}S_E\left(E\right). \tag{3.103}$$

Because of the δ-function, in the exponent we can replace $\vec{\Omega}$ with $\vec{r}/r$, thus having in the exponent the optical distance σr:

$$\Phi_0\left(\vec{r},\vec{\Omega},E\right)=\frac{\exp\left[-\sigma\left(E\right)r\right]}{4\pi r^2}\delta\left(\vec{\Omega}-\frac{\vec{r}}{r}\right)S_E\left(E\right). \tag{3.104}$$

This simple example introduces the inverse square factor $1/r^2$ for fluence attenuation with distance from the source. This formula justifies the inverse square correction widely used in dose calculations in radiotherapy. It should be noted that the inverse square approximation is valid when the following four conditions are met: (1) the size of the source is much smaller than the distance to the point of interest, then the source can be approximated by a point source; (2) the angular distribution of the source is isotropic; (3) the exponential factor can be neglected, that is, $\sigma r\ll 1$; and (4) scattered fluence can be neglected.

3.9.2 The Boltzmann Equation in Planar Geometry

If the source function and all cross sections depend on only one spatial variable, z for example, and are independent of x and y, and the medium is infinite in the (x,y) plane, then the fluence is also independent of x and y. This simplifies the streaming operator:

$$\left(\vec{\Omega}\cdot\nabla\right)\Phi\left(z,\vec{\Omega},E\right)=\mu\frac{\partial}{\partial z}\Phi\left(z,\vec{\Omega},E\right), \tag{3.105}$$

where $\mu=\Omega_z=\cos\theta$, and θ is the polar angle, that is, the angle between $\vec{\Omega}$ and the z-axis. It is usually assumed that the source and the differential scattering cross sections have azimuthal symmetry. Then the fluence is a function of only one angular variable, μ, and does not depend on the azimuthal angle ϕ. The scattering

cross section under these approximations can be written as

$$\sigma_s\left(\vec{r}, \vec{\Omega}\,' \to \vec{\Omega}, E' \to E\right) = \sigma_s\left(z, \vec{\Omega}\,' \cdot \vec{\Omega}, E' \to E\right), \tag{3.106}$$

where $\vec{\Omega}\,' \cdot \vec{\Omega}$ is the cosine of the scattering angle, which can be expressed in terms of the angular variables of the particle before, (μ', ϕ'), and after, (μ, ϕ), scattering:

$$\vec{\Omega}\,' \cdot \vec{\Omega} = \mu'\mu + \sqrt{\left(1-\mu'^2\right)\left(1-\mu^2\right)}\cos\left(\phi' - \phi\right). \tag{3.107}$$

Taking the above comments into consideration, the Boltzmann equation in planar geometry can be written as follows:

$$\begin{aligned}\mu\frac{\partial}{\partial z}\Phi\left(z,\mu,E\right) &+ \sigma\left(z,E\right)\Phi\left(z,\mu,E\right)\\ &= S\left(z,\mu,E\right) + \int \mathrm{d}\vec{\Omega}\,' \int \mathrm{d}E'\sigma_s\left(z, \vec{\Omega}\,' \cdot \vec{\Omega}, E' \to E\right)\Phi\left(z,\mu',E'\right).\end{aligned} \tag{3.108}$$

Thus, planar geometry offers a significant simplification, decreasing the dimensionality of the problem from six variables $(\vec{r}, \vec{\Omega}, E)$ to three (z, μ, E). Similar improvements can be achieved in problems with spherical and cylindrical symmetries.

3.9.3 Energy Degradation Equation

For an important class of problems, the Boltzmann equation can be simplified even further. Let us consider an infinite homogeneous medium with an infinite source with a spatially uniform distribution of activity. In this case all cross sections, the source function and fluence are independent of x, y, and z. Then, the streaming operator is zero and the Boltzmann equation, Eq. (3.40), becomes:

$$\sigma\left(E\right)\Phi\left(\vec{\Omega}, E\right) = S\left(\vec{\Omega}, E\right) + \int \mathrm{d}E' \int \mathrm{d}\vec{\Omega}\,'\sigma_s\left(\vec{\Omega}\,' \cdot \vec{\Omega}, E' \to E\right)\Phi\left(\vec{\Omega}\,', E'\right). \tag{3.109}$$

This equation can be easily integrated over $\vec{\Omega}$. The collision integral is first integrated over $\vec{\Omega}$:

$$\int \sigma_s\left(\vec{\Omega}\,' \cdot \vec{\Omega}, E' \to E\right)\mathrm{d}\vec{\Omega} = \sigma_s\left(E' \to E\right), \tag{3.110}$$

and then over $\vec{\Omega}\,'$:

$$\int \Phi\left(\vec{\Omega}\,', E'\right)\mathrm{d}\vec{\Omega}\,' = \Phi\left(E'\right). \tag{3.111}$$

Integrating the rest of Eq. (3.109) is straightforward. The resultant equation is called the energy degradation equation:

$$\sigma(E)\,\Phi(E) = S(E) + \int \mathrm{d}E'\,\sigma_s\left(E' \to E\right)\Phi\left(E'\right). \tag{3.112}$$

One important application of Eq. (3.112) is electron transport under the conditions of electron equilibrium. If electrons are produced in photon collisions, a volume V_{eq} may exist that is much larger than the electron range and thus can be considered as infinite for electrons. On the other hand, for photons the volume is small, so that photon fluence is approximately constant throughout the volume. The latter means that the source of electrons is spatially uniform throughout V_{eq}. In that case the solution, $\Phi(E)$, of Eq. (3.112) is called the electron energy degradation spectrum, equilibrium spectrum, or slowing down spectrum. Equation (3.112) is also referred to as the equilibrium equation.

3.9.4 Continuous Slowing Down Approximation

An important characteristic of charged particle transport is a very large number of inelastic collisions per unit path length. However, the energy lost in one collision is small, on the order of 10 eV. For comparison, the kinetic energy of electrons in external beam radiotherapy can be as high as a few megaelectron volts. For an electron of such a high energy, losing energy in discrete events, about 10 eV per event, is not very different from losing energy continuously. Of course, the energy spectrum of δ-electrons is very broad and it does extend to low energies. This simply means that the continuous slowing down approximation can be applied only to those electrons and other charged particles whose kinetic energy is much greater than the energy lost in one collision.

Because this approximation concerns only energy loss, we will derive it using the collision integral in the form given in the energy degradation equation, Eq. (3.112). This does not mean that the continuous slowing down approximation can be used only under equilibrium conditions. We start by noting that $\sigma_s(E' \to E)$ is a function of two variables, E', and E. These two variables are inconvenient for our derivation, so we replace one of them with the energy lost in one collision $\Delta E = E' - E$, and write the scattering cross section in the collision integral as

$$\sigma_s\left(E' \to E\right) = \sigma_s\left(E + \Delta E, \Delta E\right). \tag{3.113}$$

Function $\sigma(E', \Delta E)$ for charged particles decreases rapidly when its second argument increases. In fact, this is the reason why the energy lost in one collision tends to be small. On the other hand, variations of ΔE have a small effect on the first argument, and therefore do not cause σ_s to change significantly, because $\Delta E \ll E$. The latter arguments also apply to $\Phi(E')$. Thus we can use a Taylor series expansion:

$$\sigma(E', \Delta E)\Phi(E') \approx \sigma(E, \Delta E)\Phi(E) + \frac{\partial}{\partial E}\left[\sigma(E, \Delta E)\Phi(E)\right]\Delta E. \tag{3.114}$$

Now, we substitute the right-hand side of this equation in the collision integral:

$$\begin{aligned}
&\int_E^\infty \mathrm{d}E'\sigma_s\left(E' \to E\right)\Phi\left(E'\right) \\
&\quad \approx \int_0^\infty \sigma_s(E, \Delta E)\Phi(E)\mathrm{d}\left(\Delta E\right) + \frac{\partial}{\partial E}\int_0^\infty \left[\sigma_s(E, \Delta E)\Phi(E)\right]\left(\Delta E\right)\mathrm{d}\left(\Delta E\right) \\
&\quad = \sigma_s(E)\Phi(E) + \frac{\partial}{\partial E}\left[\beta(E)\Phi(E)\right].
\end{aligned} \tag{3.115}$$

where

$$\beta\left(E\right) = \int_0^\infty \sigma_s(E, \Delta E)\left(\Delta E\right)\mathrm{d}\left(\Delta E\right) \tag{3.116}$$

is the stopping power, that is, the average energy lost by a charged particle per unit path length

$$\beta\left(E\right) = -\frac{\mathrm{d}E}{\mathrm{d}x} \tag{3.117}$$

The units for β are MeV/cm, keV/μm, or eV/nm.

Range of a Charged Particle

Let us consider a charged particle that starts at point $x = 0$ with initial energy E_0 and travels along a straight line in the positive x direction. Knowing the rate of energy loss, $\beta(E)$, we can find the distance $x_{\max}$ the particle travels before coming to a stop at a point $x_{\max}$ where its energy, $E(x_{\max})$, is zero. To do so, we need to solve the differential equation, Eq. (3.117), with the boundary condition:

$$x\left(E_0\right) = 0. \tag{3.118}$$

The solution is straightforward. By integrating Eq. (3.117) from an arbitrary E to E_0, we have

$$x\left(E_0\right) - x\left(E\right) = -\int_E^{E_0} \frac{\mathrm{d}E'}{\beta\left(E'\right)}. \tag{3.119}$$

Then, we set $E = 0$, and note that $x(0) = x_{\max}$. Then, the range $R(E_0)$ is

$$R\left(E_0\right) = x_{\max} = \int_0^{E_0} \frac{\mathrm{d}E'}{\beta\left(E'\right)}. \tag{3.120}$$

Energy Lost Over a Finite Path Length

Problem Find the energy lost by a charged particle as it travels a distance ΔR, if its initial energy was E_1. The solution can be found using the function $R(E)$ defined by Eq. (3.120) and its inverse R^{-1}. If E_2 is the particle energy after it travels distance ΔR, then obviously

$$R\left(E_1\right) = \Delta R + R\left(E_2\right). \tag{3.121}$$

Given that $R(E)$ is a monotonic function, this equation can be solved for E_2:

$$E_2 = \begin{cases} R^{-1}\left[R\left(E_1\right) - \Delta R\right]; & \Delta R < R\left(E_1\right) \\ 0; & \Delta R \geq R\left(E_1\right) \end{cases} \tag{3.122}$$

Then, for the energy lost we have:

$$\Delta E = E_1 - E_2 = \begin{cases} E_1 - R^{-1}\left[R\left(E_1\right) - \Delta R\right]; & \Delta R < R\left(E_1\right) \\ E_1; & \Delta R \geq R\left(E_1\right) \end{cases} \tag{3.123}$$

If ΔR is so small that it can be assumed that the particle travels the distance with a constant stopping power, then

$$\Delta E = \beta\left(E_1\right) \Delta R. \tag{3.124}$$

Solution of the Energy Degradation Equation

Let us use the continuous slowing down approximation to solve the energy degradation equation, Eq. (3.112). For simplicity, we will consider a monoenergetic source, $S(E) = \delta(E - E_0)$, and assume that only one process is possible—scattering. The assumption is well justified for electrons. It means that $\sigma(E) = \sigma_s(E)$. The energy degradation equation with the collision integral written in the continuous slowing down approximation, Eq. (3.115), has the following form:

$$\sigma\left(E\right)\Phi\left(E\right) = \delta\left(E - E_0\right) + \sigma_s(E)\Phi(E) + \frac{\partial}{\partial E}\left[\beta(E)\Phi(E)\right]. \tag{3.125}$$

The collision densities $\sigma\Phi$ and $\sigma_s\Phi$ cancel out. The remaining terms are integrated over energy from an arbitrary energy E to infinity. Assuming, naturally, that fluence at infinite energy is zero, we arrive at this simple solution:

$$\Phi\left(E\right) = \begin{cases} 1/\beta\left(E\right), & E < E_0; \\ 0, & E > E_0; \end{cases} \tag{3.126}$$

Comparing this result with Eq. (3.120) leads to an important observation: the integral of fluence over energy is equal to the particle range. This finding remains valid without the continuous slowing down approximation. It follows from definition 2 of fluence (Sect. 3.1).

3.9.5 *Continuous Slowing Down Approximation for Soft Collisions*

To apply the continuous slowing down approximation, it is required that the energy lost in a collision be much less than the kinetic energy of the particle. For energetic charged particles, this requirement is met for most collisions. Still, with some probability a large energy loss in a collision is possible. This paragraph offers a more rigorous method for applying the continuous slowing down approximation. The idea is that all collisions are divided into soft and hard collisions, depending on the energy lost. A threshold energy ΔE_t is introduced, defined so that for collisions with energy loss less than ΔE_t the continuous slowing down approximation is applicable. Such collisions are called soft, and collisions with energy loss exceeding ΔE_t are called hard or catastrophic.

To demonstrate this technique, we will consider a simplified collision integral with energy variables only:

$$\int_E^\infty \sigma_s\left(E' \to E\right) \Phi\left(E'\right) \mathrm{d}E' = \int_E^{E+\Delta E_t} \sigma_s\left(E' \to E\right) \Phi\left(E'\right) \mathrm{d}E' + \int_{E+\Delta E_t}^\infty \sigma_s\left(E' \to E\right) \Phi\left(E'\right) \mathrm{d}E'. \tag{3.127}$$

The second integral accounts for collisions with energy loss exceeding ΔE_t. We will leave it unchanged. If the incident particle interacts with atomic electrons and ΔE_t is much greater than the binding energy of those electrons, modeling of such hard collisions can be simplified. For example, it may be possible to completely neglect the binding energy and simulate hard collisions as collisions with free electrons. The first integral accounts for soft collisions. We will simplify it using the continuous slowing down approximation, similar to how it was done in Eq. (3.115) except that in Eq. (3.115) integration extends to infinity but now the upper integration limit is $E + \Delta E_t$.

$$\int_E^{E+\Delta E_t} \sigma_s\left(E' \to E\right) \Phi\left(E'\right) \mathrm{d}E' = \sigma_<(E)\, \Phi(E) + \frac{\partial}{\partial E}\left[\beta_<(E)\Phi(E)\right], \tag{3.128}$$

where

$$\sigma_{<}(E) = \int_0^{\Delta E_t} \sigma_s(E, \Delta E)\, \mathrm{d}(\Delta E) \tag{3.129}$$

is the total cross section for soft collisions, and

$$\beta_{<}(E) = \int_0^{\Delta E_t} \sigma_s(E, \Delta E)\, (\Delta E)\, \mathrm{d}(\Delta E) \tag{3.130}$$

is the restricted stopping power, that is, the average energy lost in soft collisions per unit path length. The representation of the collision integral given by Eqs. (3.128)–(3.130) justifies a technique used in condensed history algorithms for charged particle transport. In this technique, each hard collision is simulated explicitly, and when a charged particle travels between hard collisions it continuously loses energy at a rate given by the restricted stopping power $\beta_{<}(E)$. Condensed history algorithms are discussed in detail in Chap. 5.

3.9.6 Fokker-Planck Approximation

This section and the next introduce two of the most basic approximations for calculating angular distributions: the Fokker-Planck and P_N approximations. The Fokker-Planck approximation is based on the assumption that scattering angles are small. It is used for charged particles in both Monte Carlo programs and grid-based Boltzmann solvers (Chap. 7). This method does not produce accurate results at large scattering angles. A good way of using the method is to apply it only to soft collisions, using another method for hard collisions as discussed in the preceding subsection.

Authors previously described the Boltzmann equation in Fokker-Planck approximation, for example, Kolchuzhkin and Uchaikin (1978) and Pomraning (1992). The derivation of the equation by Kolchuzhkin and Uchaikin (1978) is more straightforward. Therefore, we use their method.

Conceptually, the Fokker-Planck approximation is similar to the continuous slowing down approximation. It is also based on a Taylor series expansion of the collision integral but with expansion in terms of angular variables. For simplicity, we leave out energy variables, although some researchers consider the continuous slowing down approximation and the next term or terms of the expansion in energy to be part of the Fokker-Planck model.

Collision Integral in a Small-Angle Approximation

We start by applying a small-angle approximation to the collision integral

$$\begin{aligned}
&\int \mathrm{d}\vec{\Omega}\,'\sigma_s\left(\vec{r},\vec{\Omega}\,'\cdot\vec{\Omega}\right)\Phi\left(\vec{r},\vec{\Omega}\,'\right)\\
&\quad=\int_0^{2\pi}\mathrm{d}\phi\,'\int_0^{\pi}\sin\theta\,'\mathrm{d}\theta\,'\sigma_s\left(\vec{r},\vec{\Omega}\,'\cdot\vec{\Omega}\right)\Phi\left(\vec{r},\theta\,',\phi\,'\right) \qquad (3.131)\\
&\quad\approx\int_0^{2\pi}\mathrm{d}\phi\,'\int_0^{\infty}\theta\,'\mathrm{d}\theta\,'\sigma_s\left(\vec{r},\vec{\Omega}\,'\cdot\vec{\Omega}\right)\Phi\left(\vec{r},\theta\,',\phi\,'\right),
\end{aligned}$$

where we extend the upper integration limit in $\theta\,'$ from π to ∞, because in the small angle approximation, scattering to large angles close to π can be neglected. We further note that

$$\begin{aligned}
\vec{\Omega}\,'\cdot\vec{\Omega}&=\cos\theta\cos\theta\,'+\sin\theta\sin\theta\,'\cos\left(\phi-\phi\,'\right)\\
&\approx\left(1-\frac{\theta^2}{2}\right)\left(1-\frac{\theta\,'^2}{2}\right)+\theta\theta\,'\cos\left(\phi-\phi\,'\right) \qquad (3.132)\\
&\approx 1-\frac{\theta^2}{2}-\frac{\theta\,'^2}{2}+\theta\theta\,'\cos\left(\phi-\phi\,'\right),
\end{aligned}$$

and, on the other hand, that $\vec{\Omega}\,'\cdot\vec{\Omega}=\cos\theta_0\approx 1-\theta_0^2/2$, where θ_0 is the scattering angle. Using this in Eq. (3.132), we find that

$$\theta_0\approx\sqrt{\theta^2+\theta\,'^2-2\theta\theta\,'\cos\left(\phi-\phi\,'\right)}. \qquad (3.133)$$

Taylor Series Expansion of the Collision Integral

Let $\vec{u}$ be a two-dimensional vector of length θ. Its two components are $u_x=\theta\cos\phi$ and $u_y=\theta\sin\phi$. Thus, we can see in Eq. (3.133) that $\theta_0\approx\left|\vec{u}\,'-\vec{u}\,\right|$.

The last integral in Eq. (3.131) can be considered as an integral in polar coordinates, in which $\theta\,'$ is the radial variable. We then can rewrite the integral in terms of the Cartesian variables u_x and u_y;

$$\begin{aligned}
&\int_0^{2\pi}\mathrm{d}\phi\,'\int_0^{\infty}\theta\,'\mathrm{d}\theta\,'\sigma_s\left(\vec{r},\vec{\Omega}\,'\cdot\vec{\Omega}\right)\Phi\left(\vec{r},\theta\,',\phi\,'\right)\\
&\quad=\int_{-\infty}^{\infty}\mathrm{d}u_x'\int_{-\infty}^{\infty}\mathrm{d}u_y'\,\sigma_s\left(\vec{r},\left|\vec{u}\,'-\vec{u}\,\right|\right)\Phi\left(\vec{r},\vec{u}\,'\right).
\end{aligned} \qquad (3.134)$$

The substitution $\vec{w}=\vec{u}\,'-\vec{u}$ brings the integral to a form convenient for a Taylor series expansion:

$$\int_{-\infty}^{\infty} \mathrm{d}w_x \int_{-\infty}^{\infty} \mathrm{d}w_y \, \sigma_s\left(\vec{r}, \left|\vec{w}\right|\right) \Phi\left(\vec{r}, \vec{u} + \vec{w}\right). \tag{3.135}$$

In the series we retain all of the terms up to quadratic terms:

$$\begin{aligned}\Phi\left(\vec{r}, \vec{u} + \vec{w}\right) \approx \Phi\left(\vec{r}, \vec{u}\right) + \frac{\partial \Phi\left(\vec{r}, \vec{u}\right)}{\partial u_x} w_x + \frac{\partial \Phi\left(\vec{r}, \vec{u}\right)}{\partial u_y} w_y + \frac{\partial^2 \Phi\left(\vec{r}, \vec{u}\right)}{\partial u_x \partial u_y} w_x w_y \\ + \frac{\partial^2 \Phi\left(\vec{r}, \vec{u}\right)}{\partial u_x^2} \frac{w_x^2}{2} + \frac{\partial^2 \Phi\left(\vec{r}, \vec{u}\right)}{\partial u_y^2} \frac{w_y^2}{2}.\end{aligned} \tag{3.136}$$

Next, we insert Eq. (3.136) in the collision integral, Eq. (3.135). The last three terms in the first line of Eq. (3.136) are linear in w_x, w_y or both ($w_x w_y$). Therefore, the integral of these three terms is zero. We will integrate the remaining three terms one at time. The first integral produces the total scattering cross section, $\sigma_s\left(\vec{r}\right)$:

$$\int_{-\infty}^{\infty} \mathrm{d}w_x \int_{-\infty}^{\infty} \mathrm{d}w_y \, \sigma_s\left(\vec{r}, \left|\vec{w}\right|\right) \Phi\left(\vec{r}, \vec{u}\right) = \sigma_s\left(\vec{r}\right) \Phi\left(\vec{r}, \vec{u}\right) \tag{3.137}$$

The second integral is

$$\begin{aligned}&\frac{1}{2} \frac{\partial^2 \Phi\left(\vec{r}, \vec{u}\right)}{\partial u_x^2} \int_{-\infty}^{\infty} \mathrm{d}w_x \int_{-\infty}^{\infty} \mathrm{d}w_y \, \sigma_s\left(\vec{r}, \left|\vec{w}\right|\right) w_x^2 \\ &= \frac{1}{2} \frac{\partial^2 \Phi\left(\vec{r}, \vec{u}\right)}{\partial u_x^2} \int_0^{2\pi} \mathrm{d}\phi \int_0^{\infty} w \mathrm{d}w \, \sigma_s\left(\vec{r}, w\right) w^2 \cos^2\phi \\ &= \frac{1}{4} \frac{\partial^2 \Phi\left(\vec{r}, \vec{u}\right)}{\partial u_x^2} \sigma_s\left(\vec{r}\right) \langle\theta_0^2\rangle,\end{aligned} \tag{3.138}$$

where $\langle\theta_0^2\rangle$ is the mean square scattering angle. Integration of the last term is very similar

$$\begin{aligned}&\frac{1}{2} \frac{\partial^2 \Phi\left(\vec{r}, \vec{u}\right)}{\partial u_y^2} \int_{-\infty}^{\infty} \mathrm{d}w_x \int_{-\infty}^{\infty} \mathrm{d}w_y \, \sigma_s\left(\vec{r}, \left|\vec{w}\right|\right) w_y^2 \\ &= \frac{1}{4} \frac{\partial^2 \Phi\left(\vec{r}, \vec{u}\right)}{\partial u_y^2} \sigma_s\left(\vec{r}\right) \langle\theta_0^2\rangle,\end{aligned} \tag{3.139}$$

As a side note, we point out that in the small scattering angle approximation, the product $\sigma_s \langle\theta_0^2\rangle$ is closely related to the transport cross section, σ_{tr}:

$$\begin{aligned}\sigma_{\mathrm{tr}}\left(\vec{r}\right) &= \int_0^{2\pi} \mathrm{d}\phi \int_0^{\infty} w\mathrm{d}w\, \sigma_s\left(\vec{r}, w\right)\left(1-\cos w\right) \\ &\approx \int_0^{2\pi} \mathrm{d}\phi \int_0^{\infty} w\mathrm{d}w\, \sigma_s\left(\vec{r}, w\right)\frac{w^2}{2} = \frac{1}{2}\sigma_s\left(\vec{r}\right)\langle\theta_0^2\rangle,\end{aligned} \tag{3.140}$$

where we used $\cos w \approx 1-w^2/2$. In some publications, the same quantity is referred to as the momentum transfer coefficient and is denoted as α.

Collision Integral in Fokker-Planck Approximation

It follows then that in the Fokker-Planck approximation, the collision integral is replaced by a differential operator:

$$\begin{aligned}&\int \mathrm{d}\vec{\Omega}'\sigma_s\left(\vec{r}, \vec{\Omega}'\cdot\vec{\Omega}\right)\Phi\left(\vec{r}, \vec{\Omega}'\right) \\ &\quad\approx \sigma_s\left(\vec{r}\right)\Phi\left(\vec{r}, \vec{u}\right) + \frac{1}{4}\sigma_s\left(\vec{r}\right)\langle\theta_0^2\rangle\left(\frac{\partial^2}{\partial u_x^2}+\frac{\partial^2}{\partial u_y^2}\right)\Phi\left(\vec{r}, \vec{u}\right).\end{aligned} \tag{3.141}$$

The differential operator on the right-hand side of Eq. (3.141) is the Laplacian in two dimensions. It can be expressed in polar coordinates (Arfken et al. 2013), with θ as the radial variable. This yields another form of the Fokker-Planck approximation:

$$\begin{aligned}&\int \mathrm{d}\vec{\Omega}'\sigma_s\left(\vec{r}, \vec{\Omega}'\cdot\vec{\Omega}\right)\Phi\left(\vec{r}, \vec{\Omega}'\right) \\ &\quad\approx \sigma_s\left(\vec{r}\right)\Phi\left(\vec{r}, \theta, \phi\right) + \frac{1}{4}\sigma_s\left(\vec{r}\right)\langle\theta_0^2\rangle\left[\frac{1}{\theta}\frac{\partial}{\partial\theta}\left(\theta\frac{\partial}{\partial\theta}\right)+\frac{1}{\theta^2}\frac{\partial^2}{\partial\phi^2}\right]\Phi\left(\vec{r}, \theta, \phi\right)\end{aligned} \tag{3.142}$$

Finally, the above result can also be written in terms of $\mu = \cos\theta$ instead of θ. To do so, in Eq. (3.142) we replace $\theta\partial\theta$ with $\sin\theta\partial\theta = -\partial\mu$, which is the approximation we made in Eq. (3.131), and then, similarly, use $\theta^2 \approx \sin^2\theta = 1-\mu^2$. Thus, we obtain

$$\begin{aligned}&\int \mathrm{d}\vec{\Omega}'\sigma_s\left(\vec{r}, \vec{\Omega}'\cdot\vec{\Omega}\right)\Phi\left(\vec{r}, \vec{\Omega}'\right) \\ &\quad\approx \sigma_s\left(\vec{r}\right)\Phi\left(\vec{r}, \mu, \phi\right) + \frac{1}{4}\sigma_s\left(\vec{r}\right)\langle\theta_0^2\rangle\left[\frac{\partial}{\partial\mu}\left(1-\mu^2\right)\frac{\partial}{\partial\mu}+\frac{1}{1-\mu^2}\frac{\partial^2}{\partial\phi^2}\right]\Phi\left(\vec{r}, \mu, \phi\right).\end{aligned} \tag{3.143}$$

3.9.7 P_N Approximation in Planar Geometry

The Fokker-Planck approximation is based on the assumption of small scattering angles, and is normally used for charged particles, because in this case scattering is very forward-peaked. The P_N approximation works best in the opposite situation, when the angular distribution of scattered particles is broad. The best case for P_N approximation would be isotropic scattering. Let us start with the Boltzmann equation in planar geometry, Eq. (3.108), with the collision integral written in terms of the angular variables ϕ and $\mu = \cos\theta$. For brevity we will leave out energy variables:

$$\mu \frac{\partial}{\partial z} \Phi(z,\mu) + \sigma(z)\,\Phi(z,\mu) = S(z,\mu) + \int_0^{2\pi} \mathrm{d}\phi' \int_{-1}^{1} \mathrm{d}\mu' \sigma_s(z,\mu_0)\,\Phi\left(z,\mu'\right). \tag{3.144}$$

where $\mu_0 = \vec{\Omega}' \cdot \vec{\Omega}$ is cosine of the scattering angle. Then, we expand the angular dependence of fluence into Legendre series:

$$\Phi(z,\mu) = \sum_{k=0}^{\infty} \Phi_k(z)\,P_k(\mu), \tag{3.145}$$

where P_k are the Legendre polynomials, and

$$\Phi_k = \frac{2k+1}{2} \int_{-1}^{1} P_k(\mu)\,\Phi(\mu)\,\mathrm{d}\mu. \tag{3.146}$$

Similarly:

$$\sigma_s(z,\mu_0) = \sum_{n=0}^{\infty} \sigma_n(z)\,P_n(\mu_0), \tag{3.147}$$

and

$$S(z,\mu) = \sum_{n=0}^{\infty} S_n(z)\,P_n(\mu). \tag{3.148}$$

As a side note, we point out that for isotropic scattering, σ_s does not depend on μ_0, and it can be shown that all Legendre coefficients σ_n are zero, except σ_0.

Next, we substitute Eqs. (3.146) and (3.147) in the collision integral and transform it using the addition theorem:

$$P_n(\mu_0) = P_n(\mu)\,P_n\left(\mu'\right) + 2\sum_{m=1}^{n} \frac{(n-m)!}{(n+m)!} P_n^m(\mu)\,P_n^m\left(\mu'\right) \cos\left[m\left(\phi-\phi'\right)\right], \tag{3.149}$$

where P_n^m are the associated Legendre polynomials. Equation (3.149) is a form of the perhaps better known addition theorem for spherical harmonics (Arfken et al. 2013), where the spherical harmonics are written explicitly in terms of the associated Legendre polynomials. The sum in the above equation will be hereafter omitted, because the integral of $\cos[m(\phi - \phi')]$ over ϕ' is zero for all integer nonzero m. The collision integral is transformed as follows:

$$\begin{aligned}
&\int_0^{2\pi} \mathrm{d}\phi' \int_{-1}^{1} \mathrm{d}\mu' \sigma_s(z, \mu_0)\, \Phi(z, \mu') \\
&\quad = \int_0^{2\pi} \mathrm{d}\phi' \int_{-1}^{1} \mathrm{d}\mu' \sum_{n=0}^{\infty} \sigma_n(z) P_n(\mu_0) \sum_{k=0}^{\infty} \Phi_k(z) P_k(\mu') \\
&\quad = \int_0^{2\pi} \mathrm{d}\phi' \int_{-1}^{1} \mathrm{d}\mu' \sum_{n=0}^{\infty} \sigma_n(z) P_n(\mu) P_n(\mu') \sum_{k=0}^{\infty} \Phi_k(z) P_k(\mu').
\end{aligned} \tag{3.150}$$

Next, we integrate over μ' and use the orthogonality of the polynomials:

$$\int_{-1}^{1} \mathrm{d}\mu' P_n(\mu') P_k(\mu') = \frac{2}{2k+1} \delta_{nk}. \tag{3.151}$$

Then, δ_{nk} eliminates the sum over k, and integration over ϕ' produces a factor 2π. This brings us to a simplified form of the collision integral:

$$\sum_{n=0}^{\infty} \frac{4\pi}{2n+1} \sigma_n(z) P_n(\mu) \Phi_n(z). \tag{3.152}$$

After all the above modifications, Eq. (3.144) becomes:

$$\begin{aligned}
&\sum_{n=0}^{\infty} P_n(\mu) \left[\mu \frac{\partial}{\partial z} \Phi_n(z) + \sigma(z) \Phi_n(z) \right] \\
&\quad = \sum_{n=0}^{\infty} P_n(\mu) \left[S_n(z) + \frac{4\pi}{2n+1} \sigma_n(z) \Phi_n(z) \right].
\end{aligned} \tag{3.153}$$

The idea of the P_N method is as follows. The above equation is transformed into a system of $N+1$ equations for fluence coefficients Φ_k. To derive the first equation of the system, both sides of the above equation are multiplied by $P_0(\mu)$ and integrated over μ, from -1 to 1. The second equation is derived in the same fashion, using $P_1(\mu)$, and so on. For the last equation, $P_N(\mu)$ is used. For all terms, except the streaming operator, the integration is simple owing to the orthogonality of Legendre polynomials. In the streaming operator, integral

$$I_{kn} = \int_{-1}^{1} \mathrm{d}\mu \, P_k(\mu) \, \mu \, P_n(\mu) \,. \tag{3.154}$$

needs to be calculated. This integral is zero, except for two cases: case 1, where $n = k + 1$, and

$$I_{kn} = \frac{2(k+1)}{(2k+1)(2k+3)}, \tag{3.155}$$

and case 2, where $n = k - 1 \geq 0$, and

$$I_{kn} = \frac{2k}{(2k-1)(2k+1)}. \tag{3.156}$$

As a result, when Eq. (3.153) is multiplied by $P_k(\mu)$ and integrated over μ, the resultant equation includes Φ_{k-1} (if $k \geq 1$), Φ_k and Φ_{k+1}. It follows then that the last equation, where $k = N$, includes Φ_{N+1}. This makes the number of unknowns $(\Phi_0, \Phi_1, \ldots, \Phi_{N+1})$ greater than the number of equations, $N + 1$. For this reason, to make a unique solution possible, in the last equation we set Φ_{N+1} to zero.

P_1 Approximation

We will now demonstrate this method for the simple case of $N = 1$. In this case, we have a system of two equations. To derive the first equation, we multiply Eq. (3.153) by $P_0(\mu) = 1$ and integrate it over μ. In this case ($k = 0$), from the streaming operator we only have one nonzero term:

$$I_{01} = \frac{2}{3}, \tag{3.157}$$

and the first equation is:

$$\frac{1}{3}\frac{\partial}{\partial z}\Phi_1(z) + \sigma(z)\,\Phi_0(z) = S_0(z) + 4\pi\sigma_0(z)\,\Phi_0(z)\,. \tag{3.158}$$

Similarly, to derive the second equation, we multiply Eq. (3.153) by $P_1(\mu) = \mu$, and integrate over μ. In this case ($k = 1$), the streaming operator produces two nonzero terms:

$$I_{12} = \frac{4}{15}, \quad \text{and} \quad I_{10} = \frac{2}{3}. \tag{3.159}$$

We can, however, leave out I_{12}, because it is associated with Φ_2, which in P_1 approximation we must set to zero. The second equation is then as follows:

$$\frac{\partial}{\partial z}\Phi_0(z) + \sigma(z)\,\Phi_1(z) = S_1(z) + \frac{4\pi}{3}\sigma_1(z)\,\Phi_1(z)\,. \tag{3.160}$$

We now have a system of two first-order linear differential equations for the coefficients Φ_0, and Φ_1, Eqs. (3.158) and (3.160). These equations are relatively simple. Once the solution is found the fluence can be calculated:

$$\Phi(z,\mu) \approx \Phi_0(z) P_0(\mu) + \Phi_1(z) P_1(\mu) = \Phi_0(z) + \Phi_1(z)\mu. \tag{3.161}$$

P_1 approximation is also known as the diffusion approximation. Only to clarify the connection of P_1 approximation to the diffusion equation, we will make a few further simplifications. We will assume that there is only one interaction type, scattering, so that $\sigma_s = \sigma$, and that the material is uniform. That is, cross sections do not depend on z. We will also assume that the source is isotropic. In that case, $S_1 = 0$. First, in Eq. (3.158) we have:

$$4\pi\sigma_0 = 4\pi \frac{1}{2}\int_{-1}^{1} \sigma_s(\mu)\,\mathrm{d}\mu = \int_0^{2\pi} \mathrm{d}\phi \int_{-1}^{1} \mathrm{d}\mu\, \sigma_s(\mu) = \sigma_s. \tag{3.162}$$

This means that the collision integral in Eq. (3.158), $4\pi\sigma_0\Phi_0$, cancels out with the removal operator $\sigma\Phi_0$, and we have

$$\frac{1}{3}\frac{\partial}{\partial z}\Phi_1(z) = S_0(z). \tag{3.163}$$

Similarly, in Eq. (3.160) we have

$$\frac{4\pi}{3}\sigma_1 = \frac{4\pi}{3}\frac{3}{2}\int_{-1}^{1} \mu\,\sigma_s(\mu)\,\mathrm{d}\mu = \int_0^{2\pi} \mathrm{d}\phi \int_{-1}^{1} \mathrm{d}\mu\,\mu\,\sigma_s(\mu). \tag{3.164}$$

This result combined with $\sigma\Phi_1$ on the left-hand side of Eq. (3.160) produces the transport cross section:

$$\begin{aligned} &\sigma\Phi_1(z) - \int_0^{2\pi} \mathrm{d}\phi \int_{-1}^{1} \mathrm{d}\mu\ \mu\sigma_s(\mu)\,\Phi_1(z) \\ &\quad = \int_0^{2\pi} \mathrm{d}\phi \int_{-1}^{1} \mathrm{d}\mu\,(1-\mu)\,\sigma_s(\mu)\,\Phi_1(z) = \sigma_{\mathrm{tr}}\Phi_1(z). \end{aligned} \tag{3.165}$$

Recalling that for an isotropic source, $S_1 = 0$, we write Eq. (3.160) as follows:

$$\frac{\partial}{\partial z}\Phi_0(z) + \sigma_{\mathrm{tr}}\Phi_1(z) = 0. \tag{3.166}$$

Finally, from the two equations, Eqs. (3.163) and (3.166), we derive an equation for Φ_0. By taking derivative $\partial/\partial z$ of Eq. (3.166) we find $\partial\Phi_1/\partial z$, and insert this derivative in Eq. (3.163). Then, the final equation is:

$$\frac{1}{3\sigma_{\mathrm{tr}}}\frac{\partial^2\Phi_0(z)}{\partial z^2}+S_0(z)=0. \tag{3.167}$$

It formally coincides with the steady-state diffusion equation, with the diffusion coefficient $1/(3\sigma_{\mathrm{tr}})$.

3.10 Fredholm Integral Equation of the Second Kind

The Boltzmann equation can also be written in the form of an integral equation, and it will be so derived in the next section. At this point, we will only identify the type of the equation as the Fredholm integral equation of the second kind. In this section, this type of equation and solution methods are presented in a general form, without introducing any elements specific to radiation transport. Our intention is to make the material of this section applicable to a broader class of problems than just radiation transport.

The Fredholm equation of the second kind has the following general form:

$$\phi(x)=\int k\left(x'\to x\right)\phi\left(x'\right)\mathrm{d}x'+f(x), \tag{3.168}$$

where x generally is a multidimensional variable, function $f(x)$ and the kernel $k(x'\to x)$ are known, and $\phi(x)$ is unknown. For brevity it is sometimes convenient to write the equation in an operator form:

$$\phi=\hat{K}\phi+f. \tag{3.169}$$

Our purpose is to calculate a linear functional of the solution ϕ:

$$J=\int\phi(x)\,h(x)\,\mathrm{d}x=(\phi,h), \tag{3.170}$$

where $h(x)$ is a given function. The standard method for solving the equation is the following iterative algorithm:

$$\phi_i=\hat{K}\phi_{i-1}+f;\quad i=1,2,\ldots;\quad \phi_0=f; \tag{3.171}$$

which is equivalent to calculating the sum:

$$\phi=\sum_{i=0}^{\infty}\hat{K}^i f. \tag{3.172}$$

referred to as the Neumann series.

The Monte Carlo algorithm for solving Eq. (3.168) is ultimately based on the Neumann series. This will be evident from the proof of the algorithm. But first, we need to introduce the algorithm.

Algorithm Groundwork

Choose an arbitrary Markov process. It is defined by the distribution of the initial state $\pi(x)$ and transition probability $p(x' \to x)$. Both functions must be nonnegative and normalized:

$$\begin{aligned} &\int \pi(x)\,\mathrm{d}x = 1; \\ &\int p(x' \to x)\,\mathrm{d}x = p_s(x'); \quad 0 \le p_s(x') \le 1, \end{aligned} \tag{3.173}$$

where $p_s(x')$ is the "survival probability," that is, the probability that the Markov chain is not terminated at state x', and the system transitions to the next state. An additional requirement these functions must satisfy is that for all possible x and x' both ratios

$$\frac{f(x)}{\pi(x)} \quad \text{and} \quad \frac{k(x' \to x)}{p(x' \to x)} \tag{3.174}$$

are finite.

Algorithm

1. Generate a Markov chain: $\{x_0, x_1, \ldots, x_v\}$ by sampling x_0 from distribution $\pi(x_0)$, and sampling the next state from distribution $p(x' \to x)$. Before moving to the next state, a random number is drawn to determine whether or not the chain is terminated at its current state, x'. The probability of the chain terminated at x' is $1 - p_s(x')$.
2. Calculate multiplicative weights for each step:

$$Q_0 = \frac{f(x_0)}{\pi(x_0)}; Q_1 = Q_0 \cdot \frac{k(x_0 \to x_1)}{p(x_0 \to x_1)}; \cdots Q_v = Q_{v-1} \cdot \frac{k(x_{v-1} \to x_v)}{p(x_{v-1} \to x_v)}. \tag{3.175}$$

3. Calculate the sum:

$$\xi = \sum_{n=0}^{v} Q_n h(x_n), \tag{3.176}$$

where x_v is the last state of the chain.

4. Repeat (1)–(3), generate a sample $\{\xi_1, \xi_2, \ldots, \xi_N\}$. The sample size N should be large enough so that the expectation $\mathrm{E}\{\xi\}$ could be estimated within acceptably small uncertainties.
5. Estimate the functional:

$$J = (\phi, h) = \mathrm{E}\{\xi\} \approx \frac{1}{N} \sum_{i=1}^{N} \xi_i. \tag{3.177}$$

For this algorithm to work, it is required that operator $\hat{K}$ is small in a certain sense. In radiation transport, this means that particles must disappear from the system, for example, through absorption.

Proof

We now prove that the algorithm produces an unbiased estimate of the functional (ϕ, h):

$$\mathrm{E}\{\xi\} = \mathrm{E}\left\{\sum_{n=0}^{\nu} Q_n h(x_n)\right\} = (\phi, h). \tag{3.178}$$

A rigorous proof was given by Ermakov (1978). Our proof is much simplified. First, we note that the sum in the above equation has a random number of terms because the length $\nu + 1$ of the Markov chain is random. This complicates calculation of the expectation value, because the expectation of such a sum is not equal to the sum of expectations of the individual terms in the sum. We therefore extend the upper limit of the sum to infinity by introducing a factor Δ_n, such that $\Delta_n = 1$, if $n \leq \nu$ and $\Delta_n = 0$, otherwise. Then,

$$\mathrm{E}\left\{\sum_{n=0}^{\nu} Q_n h(x_n)\right\} = \mathrm{E}\left\{\sum_{n=0}^{\infty} \Delta_n Q_n h(x_n)\right\} = \sum_{n=0}^{\infty} \mathrm{E}\{\Delta_n Q_n h(x_n)\}. \tag{3.179}$$

To calculate the expectation of an individual term in the sum, we use the total probability formula:

$$\begin{aligned} &\mathrm{E}\{\Delta_n Q_n h(x_n)\} \\ &\quad = \int \mathrm{d}x_0 \int \mathrm{d}x_1 \ldots \int \mathrm{d}x_n \, p(x_0, x_1, \ldots, x_n) \, \mathrm{E}\{\Delta_n Q_n h(x_n) \,|x_0, x_1, \ldots, x_n\}. \end{aligned} \tag{3.180}$$

Here, $p(x_0, x_1, \ldots, x_n)$ is the joint probability density of the first $n + 1$ states of the Markov chain. It is a conditional distribution, the condition being that the

chain length is at least n. For this reason, in the following expression transition probabilities are divided by the respective probabilities of survival p_s:

$$p\left(x_0, x_1, \ldots, x_n\right) = \pi\left(x_0\right) \prod_{i=1}^{n} \frac{p\left(x_{i-1} \to x_i\right)}{p_s\left(x_{i-1}\right)}. \tag{3.181}$$

It can also be verified that $p(x_0, x_1, \ldots, x_n)$ given above is correctly normalized, that is, the integral of p over $x_0, x_1, \ldots, x_n$ is equal to 1. As for the conditional expectation value, we have

$$\mathrm{E}\left\{\Delta_n Q_n h\left(x_n\right) | x_0, x_1, \ldots, x_n\right\} = Q_n h\left(x_n\right) \mathrm{E}\left\{\Delta_n | x_0, x_1, \ldots, x_n\right\}. \tag{3.182}$$

Given that Δ_n is either one or zero, its expectation is equal to the probability of $\Delta_n = 1$. The latter is the probability of the system surviving all transitions starting from the initial state x_0 and reaching x_n: $x_0 \to x_1, \ldots, x_{n-1} \to x_n$:

$$\mathrm{E}\left\{\Delta_n | x_0, x_1, \ldots, x_n\right\} = \prod_{k=0}^{n-1} p_s\left(x_k\right). \tag{3.183}$$

Now we can insert Eqs. (3.181)–(3.183) and the definition of Q_n, Eq. (3.175), into the integral in Eq. (3.180):

$$\begin{aligned}
\mathrm{E}\left\{\Delta_n Q_n h\left(x_n\right)\right\} &= \int \mathrm{d}x_0 \int \mathrm{d}x_1 \ldots \int \mathrm{d}x_n \left[\pi\left(x_0\right) \prod_{i=1}^{n} \frac{p\left(x_{i-1} \to x_i\right)}{p_s\left(x_{i-1}\right)}\right] \\
&\quad \times \left[\frac{f\left(x_0\right)}{\pi\left(x_0\right)} \prod_{j=1}^{n} \frac{k\left(x_{j-1} \to x_j\right)}{p\left(x_{j-1} \to x_j\right)}\right] h\left(x_n\right) \left[\prod_{k=0}^{n-1} p_s\left(x_k\right)\right] \\
&= \int \mathrm{d}x_0 \int \mathrm{d}x_1 \ldots \int \mathrm{d}x_n f\left(x_0\right) \left[\prod_{j=1}^{n} k\left(x_{j-1} \to x_j\right)\right] h\left(x_n\right) \\
&= \left(\hat{K}^n f, h\right).
\end{aligned} \tag{3.184}$$

To verify that the very last step in the above derivation is indeed correct, the last line in the above equation may be written for a few small values of n: $n = 0$, $n = 1$, etc. We also notice that $\hat{K}^n f$ is a term in the Neumann series, Eq. (3.172). Then,

$$\mathrm{E}\left\{\xi\right\} = \sum_{n=0}^{\infty} \mathrm{E}\left\{\Delta_n Q_n h\left(x_n\right)\right\} = \sum_{n=0}^{\infty} \left(\hat{K}^n f, h\right) = (\phi, h). \tag{3.185}$$

This proves that ξ defined by Eqs. (3.175)–(3.176) is indeed an unbiased estimate of the functional (ϕ, h) of solution ϕ of the Fredholm integral equation of the second kind. □

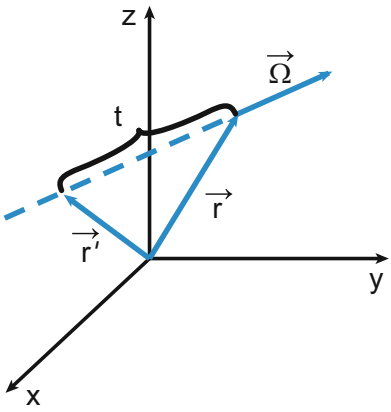

Fig. 3.7 Derivation of the Boltzmann equation in an integral form

3.11 The Boltzmann Equation in an Integral Form

The Boltzmann equation can be written in an integral form. Its derivation can be performed formally by integrating the integro-differential form of the equation, Eq. (3.40), along a straight line leading to a point $\vec{r}$ from the boundary of the computational domain Γ, in direction $\vec{\Omega}$. We, instead, offer a more heuristic derivation that provides a better insight into the underlying physics.

For brevity, we will exclude particle energy from consideration, until the very last equation. Fluence $\Phi(\vec{r}, \vec{\Omega})$ at a point $\vec{r}$ we represent as a sum of unscattered fluence $\Phi_0(\vec{r}, \vec{\Omega})$ and fluence of particles that after scattering at a point $\vec{r}\,'$ reach $\vec{r}$ without any more scattering, Fig. 3.7. Integration over all possible locations $\vec{r}\,'$ is implied.

Next, we note that when calculating fluence $\Phi(\vec{r}, \vec{\Omega})$ we need to count only those particles that travel in direction $\vec{\Omega}$ when they reach $\vec{r}$. This means that when integrating over $\vec{r}\,'$ we only need to integrate along the dashed line in Fig. 3.7. That is, we integrate from $\vec{r}$ to the boundary of the computational domain Γ, along a straight line, in the direction opposite to $\vec{\Omega}$. The number of interactions at $\vec{r}\,'$ is given by the collision density $\sigma\Phi$, except that we count only those collisions that produce particles traveling in direction $\vec{\Omega}$. For this reason, in the expression for collision density below we use the differential cross section:

$$\sigma_s\left(\vec{r}\,', \vec{\Omega}\,' \to \vec{\Omega}\right) \Phi\left(\vec{r}\,', \vec{\Omega}\,'\right). \tag{3.186}$$

This gives us the number of particles that were traveling in direction $\vec{\Omega}\,'$ and scattered at $\vec{r}\,'$ exactly in the direction of point $\vec{r}$. To account for all $\vec{\Omega}\,' \to \vec{\Omega}$ scattering events, we will need to integrate over all $\vec{\Omega}\,'$. A fraction of these scattered

particles will reach $\vec{r}$ without any further scattering. The size of this fraction is subject to exponential attenuation and is given by this factor:

$$\exp\left[-\tau\left(\vec{r},\vec{r}\,'\right)\right], \tag{3.187}$$

where $\tau(\vec{r},\vec{r}\,')$ is the optical distance between points $\vec{r}$ and $\vec{r}\,'$.

As we have already mentioned, we need to integrate along the dashed line in Fig. 3.7. The equation of the line is

$$\vec{r}\,'(t) = \vec{r} - \vec{\Omega}t. \tag{3.188}$$

The limits of integration are from $t = 0$ to $t = t_{\max}$, where $t_{\max}$ is the distance from $\vec{r}$ to the point where the line reaches the boundary Γ of the computational domain. Adding the two components of fluence, unscattered and scattered, we arrive at the following form of the Boltzmann equation:

$$\Phi\left(\vec{r},\vec{\Omega}\right) = \Phi_0\left(\vec{r},\vec{\Omega}\right) + \int_0^{t_{\max}} \mathrm{d}t \int \mathrm{d}\vec{\Omega}\,' \exp\left[-\tau\left(\vec{r},\vec{r}-\vec{\Omega}t\right)\right]\sigma_s\left(\vec{r}-\vec{\Omega}t,\vec{\Omega}\,'\to\vec{\Omega}\right)\Phi\left(\vec{r}-\vec{\Omega}t,\vec{\Omega}\,'\right). \tag{3.189}$$

The final step is the transformation of the line integral into a volume integral. For that purpose, we use the following identity (see Appendix B):

$$\int_0^{\infty} f\left(\vec{r}-t\,\vec{\Omega}\right)\mathrm{d}t = \int f\left(\vec{r}\,'\right)\delta\left(\vec{\Omega}-\frac{\vec{r}-\vec{r}\,'}{|\vec{r}-\vec{r}\,'|}\right)\frac{\mathrm{d}\vec{r}\,'}{|\vec{r}-\vec{r}\,'|^2}. \tag{3.190}$$

In the final equation, for completeness we include the particle energy:

$$\Phi\left(\vec{r},\vec{\Omega},E\right) = \Phi_0\left(\vec{r},\vec{\Omega},E\right) + \int_V \mathrm{d}\vec{r}\,' \int \mathrm{d}\vec{\Omega}\,' \int \mathrm{d}E'\,\frac{\exp\left[-\tau\left(\vec{r},\vec{r}\,',E\right)\right]}{|\vec{r}-\vec{r}\,'|^2} \times \sigma_s\left(\vec{r}\,',\vec{\Omega}\,'\to\vec{\Omega},E'\to E\right)\Phi\left(\vec{r}\,',\vec{\Omega}\,',E'\right)\delta\left(\vec{\Omega}-\frac{\vec{r}-\vec{r}\,'}{|\vec{r}-\vec{r}\,'|}\right). \tag{3.191}$$

It should be clear from the derivation that the optical distance $\tau(\vec{r},\vec{r}\,',E)$ is calculated for energy E, which is the particle energy *after* it interacted at point $\vec{r}\,'$. It is equal to the particle energy when it arrives at point $\vec{r}$, *before* the next interaction. The result, Eq. (3.191), is a type of integral equation known as the Fredholm equation of the second kind and can be written in the following general form:

$$\Phi(x) = \Phi_0(x) + \int \Phi\left(x'\right)k\left(x'\to x\right)\mathrm{d}x'. \tag{3.192}$$

The derivation of the adjoint integral equation is similar. It has the same form as the integral equation for the fluence, except that the kernel of the integral operator is transposed:

$$\Phi^{+}(x) = \Phi_0^{+}(x) + \int k\left(x \to x'\right) \Phi^{+}\left(x'\right) \mathrm{d}x'. \tag{3.193}$$

3.12 The Boltzmann Equation as a Basis for Biasing Techniques

We have shown that the Boltzmann equation in its integral form is a Fredholm equation of the second kind. We have also introduced a Monte Carlo algorithm for solving equations of exactly that type. Thus, we now have a general Monte Carlo algorithm for solving problems of radiation transport.

To apply this general method to a specific problem, we need to choose a Markov process. We can choose a process that is a copy, or more realistically, a close model of the actual physical process, for example, propagation of X-rays in a material. The advantage of this choice is that the algorithm will be very intuitive.

On the other hand, we do not have to copy the actual process. We can pick an almost arbitrary Markov process. Choosing a process that is not an exact copy of the actual process is an optimization technique called biasing. Processes, or particle trajectories generated on a computer, are different or "biased" compared with real ones, but the final result, the solution, such as the particle fluence, is unbiased. Biasing of trajectories is exactly compensated for by weights Q_n, Eq. (3.175).

3.12.1 Source Biasing

To introduce source biasing, we start with the adjoint representation of detector reading

$$J = \left(S, \Phi^{+}\right) = \int \mathrm{d}\vec{r} \int \mathrm{d}\vec{\Omega} \int \mathrm{d}E\, S\left(\vec{r}, \vec{\Omega}, E\right) \Phi^{+}\left(\vec{r}, \vec{\Omega}, E\right). \tag{3.194}$$

The definitions of the source function and adjoint function (Sect. 3.1) offer a statistical interpretation of the above integral. In this representation, J is the expectation of $\Phi^{+}(\vec{r}, \vec{\Omega}, E)$, where the distribution of phase coordinates $(\vec{r}, \vec{\Omega}, E)$ is $S(\vec{r}, \vec{\Omega}, E)$. To calculate this integral, we would sample $(\vec{r}, \vec{\Omega}, E)$ from distribution $S(\vec{r}, \vec{\Omega}, E)$ and then calculate the contribution to detector reading from a particle that started with $(\vec{r}, \vec{\Omega}, E)$. We do not need to specify at this point how this contribution would be calculated. It is implied, of course, it would be done by generating the rest of the particle trajectory.

Now, let us introduce in Eq. (3.194) an arbitrarily biased source function, $\tilde{S}$:

$$J = \int \mathrm{d}\vec{r} \int \mathrm{d}\vec{\Omega} \int \mathrm{d}E \, \tilde{S}\left(\vec{r}, \vec{\Omega}, E\right) w\left(\vec{r}, \vec{\Omega}, E\right) \Phi^{+}\left(\vec{r}, \vec{\Omega}, E\right). \tag{3.195}$$

This introduces a variable, $w = S/\tilde{S}$, called the particle weight or more specifically its starting weight. This new representation for J has a slightly different statistical interpretation. The initial phase coordinates of the particle, $(\vec{r}, \vec{\Omega}, E)$, are now sampled from the biased distribution $\tilde{S}(\vec{r}, \vec{\Omega}, E)$, and its contribution to detector reading $\Phi^{+}(\vec{r}, \vec{\Omega}, E)$ is multiplied by the weight $w(\vec{r}, \vec{\Omega}, E)$. The weight is assigned to the particle as soon as its initial phase coordinate is known. If particle i starts with coordinates $(\vec{r}_i, \vec{\Omega}_i, E_i)$, then its weight is: $w_i = S(\vec{r}_i, \vec{\Omega}_i, E_i)/\tilde{S}(\vec{r}_i, \vec{\Omega}_i, E_i)$. The weight does not change as the particle travels and interacts, unless other, additional biasing techniques are applied. If the particle produces secondary particles, those particles inherit the weight. If, for example, the code counts the number of particles crossing a certain surface, and no biasing techniques are used, then the count increases by one when a particle crosses the surface. If source biasing is used, then the count increases by w_i.

Example 1

A large source with a spatially uniform activity, isotropically emitting particles with a polyenergetic spectrum, and a small detector at some distance from the source. Problem: too many particles never reach the detector.

To address this problem, we can bias the spatial or angular distribution of the source, or its energy spectrum. When biasing the spatial distribution, the activity of the source should be increased in the area close to the detector. Particles starting in that area will have a better chance of reaching the detector. They will have a weight $w < 1$. It is generally recommended that the biased source be similar to the original source, so that the weights do not fluctuate too much.

The angular distribution should be biased so as to increase the number of particles emitted by the source in the direction of the detector. An example is in Fig. 3.8, which shows two angular distribution functions: the original isotropic distribution (blue line) and a biased distribution (purple line). The two distributions are shown as a function of $\mu = \cos\theta$, where direction $\theta = 0$ points at the center of the detector. Particles that start at angles corresponding to the part of the purple line that is above the blue line will have a weight $w < 1$.

Biasing the energy spectrum may also be beneficial. Here are two examples where it might work.

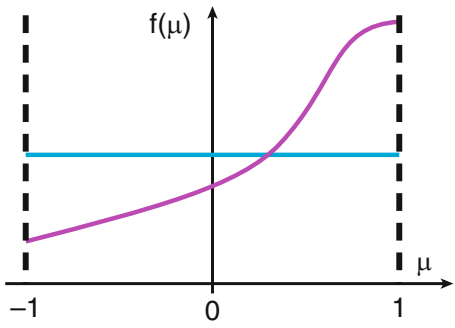

Fig. 3.8 Biasing the angular distribution of the source. Shown are the original unbiased distribution (*blue line*) and a biased distribution (*purple line*); $\mu = \cos\theta$, where $\theta = 0$ is the direction towards the detector

Example 2

Too many particles are absorbed before they can reach the detector, and we are counting particles entering the detector volume through its surface. Then, we can bias the source spectrum so as to increase the number of high-energy particles. If cross sections decrease with increasing energy, as is often the case, fewer particles will be absorbed as they travel from the source to the detector.

Example 3

We are counting ionization collisions within a detector. Most particles reach the detector, but too many cross the detector volume without interactions. Then we can increase the number of low-energy particles in the biased spectrum. If, the cross sections, again, decrease with increasing energy, this biasing will increase the collision density and therefore increase the number of ionizations within the detector. In the next chapter we present a better method for addressing this problem. In this method a particle entering a volume interacts within it with a probability defined by the user. The method is called forced interactions. It does not introduce any systematic error.

3.12.2 Trajectory Biasing

This section presents a general method for optimization of Monte Carlo algorithms for solving the Boltzmann equation. It is based on the method we introduced in Sect. 3.10 for solving Fredholm integral equations of the second kind. We will show how a correct, unbiased solution of the Boltzmann equation can be obtained with an algorithm based on simulation of an almost arbitrary Markov process. The method does not require copying the actual process of particle transport. The correct solution is obtained using particle trajectories that are biased, different in a statistical sense, from real particle trajectories. To reiterate, particle trajectories are

biased, but the solution calculated using these trajectories is unbiased. Allowing an algorithm developer to use almost any Markov process without introducing any systematic errors offers virtually unlimited flexibility in designing an optimal algorithm. The priority is certainly minimizing statistical uncertainties without increasing the computing time or introducing other sources of error. Techniques aimed at achieving these goals are called variance reduction. Many, or perhaps most, popular variance reduction techniques are ultimately just special cases of the method presented in this section. We will return to this point in the next section where we discuss in detail some of those techniques.

First, we switch from fluence to collision density: $F(x) = \sigma\Phi(x)$, where σ is the total cross section, which is a function of $\vec{r}$ and E. The integral equation for the collision density $F(x)$ can be easily derived from Eq. (3.192):

$$F(x) = F_1(x) + \int k_F(x' \to x) F(x') \, dx', \tag{3.196}$$

where, $F_1 = \sigma\Phi_0$, is called the first collision density, and

$$k_F(x' \to x) = k(x' \to x) \sigma(x) / \sigma(x'). \tag{3.197}$$

Our purpose is to calculate the functional:

$$J = (\Phi, D) = \left(F, \frac{D}{\sigma}\right), \tag{3.198}$$

where $D = D(x)$ is the detector response function defined in Sect. 3.1. To calculate the functional we use the general algorithm for solving Fredholm equations presented in Sect. 3.10. The main component of the algorithm is the unbiased estimate ξ of the functional J, defined by Eq. (3.176). In the case of the equation for the collision density it has the form:

$$\xi = \sum_{n=0}^{\nu} Q_n \frac{D(x_n)}{\sigma(x_n)}. \tag{3.199}$$

This estimate belongs to the so-called collision-type estimates. This and several other types are discussed in Sect. 4.3.

This general algorithm produces the correct solution with an almost arbitrary Markov chain. In practice, however, for solving problems of radiation transport Markov processes are used that are similar to the physical process of particle propagation in matter. Hence, the technique can be described as trajectory biasing: particle trajectories are generated, but they are different in a statistical sense from trajectories of real particles. In this section we demonstrate using a simple example how this technique is applied. More examples are given in Sect. 4.4, where variance reduction techniques are discussed.

Problem

Using real, unbiased trajectories generated in water, calculate detector reading J, for example, dose distribution in another material. In practice, a large number of trajectories in water are generated, stored on a disk, and then used as needed to calculate J in other materials. An obvious benefit of this method is that trajectories are generated only once. It works best if the other materials are not very different from water in terms of particle interaction parameters. Alternatively, particle trajectories are not saved to disk, instead, as particle trajectories in water are generated, J is calculated in water, and simultaneously in a number of other materials. This method may save computing time because there is no need to generate particle trajectories in any material other than water.

We do not bias the source here, we focus on biasing trajectories. We assume for simplicity that all materials are homogeneous.

First, we show that for unbiased trajectories the transition probability $p(x_i \to x_{i+1})$ coincides with the kernel of the integral equation $k_F(x_i \to x_{i+1})$. To do so, we derive an algorithm for sampling transition $x_i \to x_{i+1}$, when the transition probability is equal to $k_F(x_i \to x_{i+1})$.

Proof

Within the assumptions we made for this problem, it follows from Eq. (3.191) that

$$k_F\left(x_i \to x_{i+1}\right) = \sigma\left(E_{i+1}\right) \exp\left[-\sigma\left(E_{i+1}\right) \cdot \left|\vec{r}_i - \vec{r}_{i+1}\right|\right] \times \frac{\sigma_s\left(\vec{\Omega}_i \cdot \vec{\Omega}_{i+1}, E_i \to E_{i+1}\right)}{\sigma\left(E_i\right)} \delta\left(\vec{\Omega}_{i+1} - \frac{\vec{r}_{i+1} - \vec{r}_i}{\left|\vec{r}_{i+1} - \vec{r}_i\right|}\right) \frac{1}{\left|\vec{r}_{i+1} - \vec{r}_i\right|^2}. \tag{3.200}$$

We interpret the above expression as the conditional probability distribution of x_{i+1} given x_i. That is, for any given x_i we will use the above function k_F to sample x_{i+1}. The formula simplifies somewhat if instead of sampling $\vec{r}_{i+1}$ we sample $\vec{R} = \vec{r}_{i+1} - \vec{r}_i$. We will also write $\vec{R} = R\,\vec{\Omega}_R$. Then,

$$k_F(x_i \to x_{i+1}) = \sigma\left(E_{i+1}\right) \exp\left[-\sigma\left(E_{i+1}\right) R\right] \times \frac{\sigma_s\left(\vec{\Omega}_i \cdot \vec{\Omega}_{i+1}, E_i \to E_{i+1}\right)}{\sigma\left(E_i\right)} \delta\left(\vec{\Omega}_{i+1} - \vec{\Omega}_R\right) \frac{1}{R^2}. \tag{3.201}$$

The above expression is a product of probabilities. First, we consider the factor $\sigma_s(\ldots)/\sigma$. It is the distribution of $\vec{\Omega}_{i+1}$ and E_{i+1}, except it is not normalized to 1:

$$\frac{1}{\sigma\left(E_i\right)}\int \mathrm{d}\vec{\Omega}_{i+1}\int \mathrm{d}E_{i+1}\sigma_s\left(\vec{\Omega}_i\cdot\vec{\Omega}_{i+1},E_i\to E_{i+1}\right)=\frac{\sigma_s\left(E_i\right)}{\sigma\left(E_i\right)}=P_s\left(E_i\right)\le 1, \tag{3.202}$$

where P_s is the probability of a particle surviving step $x_i \to x_{i+1}$. Then, before making this step, we draw a random number γ to determine whether or not the particle survives this step. If it does, we sample its new energy, E_{i+1}, and direction, $\vec{\Omega}_{i+1}$, from the now normalized distribution:

$$f\left(\vec{\Omega}_{i+1},E_{i+1}|\vec{\Omega}_i,E_i\right)=\frac{\sigma_s\left(\vec{\Omega}_i\cdot\vec{\Omega}_{i+1},E_i\to E_{i+1}\right)}{P_s\left(E_i\right)\sigma\left(E_i\right)}. \tag{3.203}$$

Next, to sample vector $\vec{R}$, we sample its length R and direction $\vec{\Omega}_R$. To do so, we note the following relation between distribution of a vector $f(\vec{R})$ and distribution of its length, $f(R)$:

$$f\left(\vec{R}\right)\mathrm{d}\vec{R}=f\left(R\right)R^2\mathrm{d}R\mathrm{d}\vec{\Omega}_R. \tag{3.204}$$

This eliminates factor $1/R^2$ from Eq. (3.201), and we can sample R, which is the distance to the next point of the Markov chain, from the exponential distribution:

$$f\left(R\right)=\sigma\left(E_{i+1}\right)\exp\left[-\sigma\left(E_{i+1}\right)R\right]. \tag{3.205}$$

The last term in Eq. (3.201) that we have not considered yet is the delta function $\delta(\vec{\Omega}_{i+1}-\vec{\Omega}_R)$. It simply means that the direction $\vec{\Omega}_R$ of particle travel from point $\vec{r}_i$ to the next point, must be $\vec{\Omega}_{i+1}$, which we had sampled earlier from the distribution given by Eq. (3.203).

The sampling procedure that we have just described is identical to generating an unbiased particle trajectory: (1) at point $\vec{r}_i$ a type of particle interaction, absorption or scattering, is sampled; (2) if it is scattering, a new particle energy, E_{i+1}, and direction $\vec{\Omega}_{i+1}$ are sampled according to the differential scattering cross section; (3) from point $\vec{r}_i$, the particle travels in direction $\vec{\Omega}_{i+1}$; (4) a distance R to the next interaction point (free path) is sampled from the exponential distribution. □

It follows that in the above described algorithm, the ratio $k(x_i \to x_{i+1})/p(x_i \to x_{i+1})$, needed to calculate the weight Q_{i+1} in Eq. (3.175), is equal to one. In that case, to make all the weights Q_n equal to one, the initial weight Q_0 must also be equal to one. To achieve that, according to Eqs. (3.175) and (3.196), the initial phase coordinate x_0 must be sampled from distribution

$$F_1\left(x_0\right)=\sigma\left(E_0\right)\Phi_0\left(x_0\right)=\sigma\left(E_0\right)\int_0^\infty \mathrm{d}t\,S\left(\vec{r}_0-\vec{\Omega}_0t,\vec{\Omega}_0,E_0\right)\exp\left[-\sigma\left(E_0\right)t\right], \tag{3.206}$$

where we used Eq. (3.95) for the unscattered fluence Φ_0. To derive an algorithm for sampling the initial phase coordinate x_0 from the distribution given by Eq. (3.206),

we first, transform the line integral in Eq. (3.206) into a volume integral using Eq. (3.190), and then follow the method that we used above to derive the algorithm for sampling transition $x_i \rightarrow x_{i+1}$. This produces the following algorithm for sampling from distribution $F_1(x_0)$.

Algorithm

1. Sample $\vec{r}\,' = \vec{r}_0 - \vec{\Omega}_0 t$ from the source distribution, $S(\vec{r}\,')$.
2. Sample $\vec{\Omega}_0, E_0$ from the source distribution $S(\vec{r}\,', \vec{\Omega}_0, E_0)$.
3. Sample free path t from the exponential distribution $\sigma(E_0)\exp[-\sigma(E_0)t]$.
4. Calculate $\vec{r}_0 = \vec{r}\,' + \vec{\Omega}_0 t$.

This algorithm is no different than sampling the first step of an unbiased particle trajectory. We can therefore conclude that all weights Q_n are equal to one, if the algorithm copies the actual particle transport process.

In the context of the problem, unbiased trajectories would be those generated in the other material, not water. However, what we have are trajectories in water. We consider them as biased. The only difference between biased and unbiased processes are the cross sections. We mark biased cross sections with a tilde, e.g., $\tilde{\sigma}$. The transition probability for this biased process has exactly the same form as $k_F(x_i \rightarrow x_{i+1})$ in Eq. (3.200), but with biased cross sections:

$$p(x_i \rightarrow x_{i+1}) = \tilde{\sigma}(E_{i+1}) \exp\left[-\tilde{\sigma}(E_{i+1}) \cdot |\vec{r}_i - \vec{r}_{i+1}|\right] \times \frac{\tilde{\sigma_s}\left(\vec{\Omega}_i \cdot \vec{\Omega}_{i+1}, E_i \rightarrow E_{i+1}\right)}{\tilde{\sigma}(E_i)} \delta\left(\vec{\Omega}_{i+1} - \frac{\vec{r}_{i+1} - \vec{r}_i}{|\vec{r}_{i+1} - \vec{r}_i|}\right) \frac{1}{|\vec{r}_{i+1} - \vec{r}_i|^2}. \tag{3.207}$$

Then, from Eq. (3.175) the weight for transition from x_i to x_{i+1} is calculated as follows:

$$\begin{aligned} Q_{i+1} &= Q_i \frac{k_F(x_i \rightarrow x_{i+1})}{p(x_i \rightarrow x_{i+1})} \\ &= Q_i \frac{\sigma(E_{i+1})}{\tilde{\sigma}(E_{i+1})} \exp\{[\tilde{\sigma}(E_{i+1}) - \sigma(E_{i+1})] R_{i+1}\} \frac{\sigma_s\left(\vec{\Omega}_i \cdot \vec{\Omega}_{i+1}, E_i \rightarrow E_{i+1}\right)}{\tilde{\sigma_s}\left(\vec{\Omega}_i \cdot \vec{\Omega}_{i+1}, E_i \rightarrow E_{i+1}\right)} \frac{\tilde{\sigma}(E_i)}{\sigma(E_i)}. \end{aligned} \tag{3.208}$$

where R_{i+1} is the free path, that is the distance between $\vec{r}_i$ and $\vec{r}_{i+1}$. The initial weight, for $i = 0$, is

$$Q_0 = \frac{F_1(x_0)}{\tilde{F}_1(x_0)} = \frac{\sigma(E_0)}{\tilde{\sigma}(E_0)} \frac{\Phi_0(x_0)}{\tilde{\Phi}_0(x_0)}, \tag{3.209}$$

where $\Phi_0(x_0)$ and $\tilde{\Phi}_0(x_0)$ are, respectively, the unbiased and biased unscattered fluences at point $\vec{r}_0$, in direction $\vec{\Omega}_0$, from a particle that started from the source with energy E_0.

To summarize: particle trajectories $\{x_0, \ldots x_v\}$ are generated only in water. To calculate detector reading J in another material, the estimate ξ given by Eq. (3.176) is used with the weights calculated according to Eqs. (3.208) and (3.209). This produces an unbiased result, that is, $\mathrm{E}\{\xi\} = J$.

Exercise: Optimization of Shielding Calculations

A one-dimensional problem. Particles are incident normally on a one meter thick wall. The cross sections are: total $\sigma = 0.1\,\mathrm{cm}^{-1}$; scattering $\sigma_s = 0.05\,\mathrm{cm}^{-1}$; and absorption $\sigma_A = 0.05\,\mathrm{cm}^{-1}$. Assume that scattering does not change any properties of the particle, the particle continues traveling in the same direction (this is called delta scattering).

1. Calculate analytically the number of particles that exit the wall, per incident particle. In other words, find the probability of an incident particle penetrating the wall.
2. Write a Monte Carlo code that estimates the above quantity and calculates statistical uncertainty of the estimate. Sample the free path from exponential distribution with parameter σ, move the particle to the collision point, and sample the type of interaction using the cross-sectional data. If it is absorption, terminate particle history. The history is also terminated when the particle exits the wall.
3. Run the Monte Carlo code and compare the result with the analytical solution.
4. Write another Monte Carlo code that uses trajectory biasing in an attempt to reduce statistical uncertainties. Reduce the density of the wall. More specifically, reduce all cross sections by a factor δ (we call it the dilution coefficient). To compensate for the bias, the multiplicative weights Q_n need to be calculated, Eq. (3.208). Do not bias the source.
5. Run the second Monte Carlo code for several different dilution coefficients. Make sure that you always get a result consistent with your analytical solution. Each time calculate statistical uncertainties. Find a δ that minimizes the uncertainties. A graph of the standard deviation versus δ may help. Do not use very large or small δ.
6. Compare the CPU times, calculate and compare algorithm efficiencies [Eq. (4.39)].
7. Note that the weights do not need to be calculated for every particle and at every step. Instead, they can be calculated only for those particles that exit the wall and only when they do so.
8. When calculating the weights, carefully consider the last step, where the free path exceeds the distance to the wall boundary.

References

Arfken, G.B., Weber, H.J., Harris, F.E.: Mathematical Methods for Physicists, 7th edn. Elsevier, Amsterdam (2013)

Balescu, R.: Equilibrium and Non-Equilibrium Statistical Mechanics. Wiley, New York (1975)

Ermakov, S.M.: Unbiased estimates of the Neumann series sum by the Monte Carlo method. J. Sov. Math. **9**(6), 963–974 (1978)

Granger, R.A.: Fluid Mechanics. Dover, New York (1995)

International Commission on Radiation Units and Measurements: Fundamental Quantities and Units for Ionizing Radiation. ICRU Report 60 (1998)

Jackson, J.D.: Classical Electrodynamics, 3rd edn. Wiley, New York (1999)

Kolchuzhkin, A.M., Uchaikin V.V.: Introduction to the Theory of Penetration of Particles through the Matter. Atomizdat, Moscow (1978)

Landau, L.D., Lifshitz E.M.: Statistical Physics, 2nd revised and enlarged edition. Pergamon Press, Oxford (1969)

Polyakov, P.A.: Bogolyubov (BBGKY) hierarchy in classical relativistic electrodynamics. Theor. Math. Phys. **76**(3), 939–944 (1988)

Pomraning, G.C.: The Fokker-Planck operator as an asymptotic limit. Math. Models Methods Appl. Sci. **2**(1), 21–36 (1992)

Vassiliev, O.N., Wareing, T.A., McGhee, J., Failla, G., Salehpour, M.R.: Validation of a new grid-based Boltzmann equation solver for dose calculation in radiotherapy with photon beams. Phys. Med. Biol. **55**(3), 581–598 (2010)

Wienke, B.R.: Transport equation in modified Eulerian coordinates. Phys. Fluids **17**(6), 1135–1138 (1974)

Wienke, B.R.: ESN: one-dimensional *Sn* transport module for electrons. J. Quant. Spectrosc. Radiat. Transf. **28**(4), 311–326 (1982)

Chapter 4
Particle Trajectories, Tallies, and Variance Reduction

4.1 Planning Monte Carlo Calculations

Defining the Source

A source function defines probability distributions of all parameters of all emitted particles. The parameters include phase coordinates and internal degrees of freedom, of which at least the type of particles must be specified. If more than one type of particles is emitted, then a discrete distribution of particle types must be specified. A real source may emit more than one particle at a time, and the number of particles can be random. A launch of several particles at a time can be easily implemented in a Monte Carlo code. However, for most standard problems, such as calculations of dose distributions, such implementation is not necessary, and particles can be simulated one at a time. The final result, because of linearity of the Boltzmann equation, can be rescaled to account for the actual total number of particles of each type emitted by a source. The types of problems where the results cannot be simply scaled by the number of source particles are calculations of fluctuations and correlations of radiation fields, see for example Wang and Vassiliev (2014). In those cases, it may be necessary to simulate a simultaneous launch of several particles. However, such a simulation can be avoided in some cases where special "rescaling" methods have been developed. An example of the "rescaling" methods is provided later by Eq. (6.3) in Chap. 6. An additional complication for problems concerning fluctuations and correlations is that correlations between particles emitted by the source, if they are present, need to be accounted for.

Spatial, angular, and energy distributions of a source must be defined. The source can be located at a point, occupy a volume, or spread on a surface. For surface and volume sources, a spatial distribution of their activities must be specified. The simplest form of a spatial distribution is uniform. For example, for radioisotope brachytherapy sources the activities may be uniform on a seed surface or in the

O.N. Vassiliev, *Monte Carlo Methods for Radiation Transport*, Biological and Medical Physics, Biomedical Engineering, DOI 10.1007/978-3-319-44141-2_4

seed volume, depending on the design. In terms of angular distribution a source can be isotropic, mono-directional (usually for a narrow beam), or a parallel beam, either collimated or infinite in the lateral direction. The types of source energy spectra include mono-energetic, discrete, and continuous. These distributions can be specified by formulas or tables.

Parameters of a particle emitted by a source, such as its initial position $\vec{r}_0$, direction $\vec{\Omega}_0$, and energy E_0, are not necessarily statistically independent. If there is indeed a correlation between any of particle's initial parameters, methods for sampling joint distributions should be used to model the source (Sect. 2.12).

Time-Varying Systems

In some problems, properties of systems change with time during irradiation. Usually the changes are in the geometry and material properties. Important examples in radiotherapy are changes in patient anatomy caused by respiratory motion, and modulation of incident fluence achieved through a complex sequence of changes in the shape of beam aperture, beam direction, and in some techniques beam intensity. In most cases, temporal variations in an irradiated system do not pose a problem, because an individual particle traverses the system much faster than any significant changes can occur in the system. Then, particle trajectories can be generated in a static geometry representing a snapshot of the system at a certain moment of time. Accordingly, for radiation transport calculations temporal evolution of a system is approximated by a series of snapshots closely spaced in time. For each snapshot, a large number of particle trajectories are generated so as to achieve an acceptably small overall uncertainty. For example, patient dose calculations that account for respiratory motion are often performed separately for each of the ten phases of a respiratory cycle. Then, the combined dose, delivered over a course of treatment is calculated. The latter task is nontrivial and requires special methods that are beyond the scope of this book. For those problems that do require temporal information on particle trajectories, some standard Monte Carlo software can generate timestamps for particle events, such as interactions. If it is also needed to sample temporal patterns of particle production by a source, a simple model that can be used for the purpose is the Poisson process. In this model, the next particle starts after a random time interval Δt that is exponentially distributed. The average of Δt is the only parameter of the distribution, and it can be calculated from the source activity.

Defining the Geometry and Materials

When defining the geometry, two important question are, as to what objects need to be included in simulations and how large a simulation volume (i.e., the computational domain) should be. Standard boundary conditions require that trajectories of

those particles that have left the simulation volume are terminated. However, making a simulation volume too small may introduce significant systematic errors because of the loss of particles that may have a good chance of scattering towards and reaching the detector. On the other hand, making the volume too big slows down the calculations, because computing time is wasted on tracking particles that never reach the detector. If electric or magnetic fields are present, they affect transport of charged particles and therefore need to be specified. Some of the standard Monte Carlo software are capable of generating particle trajectories in electric and magnetic fields.

Defining the Physics

Planning Monte Carlo calculations requires answering several important questions concerning the physics of a particular problem under consideration. For example, which particle types need to be tracked? If the source emits photons, is it necessary to track delta electrons and positrons? And, if electrons are tracked, can bremsstrahlung photons be neglected? Some interaction types leave the atom in an excited state. In that case, is it necessary to model de-excitation processes, such as production of fluorescence photons and Auger electrons? For each particle type, the energy range should be defined. What is the maximum energy a particle can have? The maximum particle energy does not necessarily coincide with the maximum source energy. Particles can gain energy from nuclear reactions, or through acceleration by an external electric field, if it is present. As for the minimum energy, it is a standard practice to define for charged particles the so-called tracking cut-offs. When particle energy falls below the cut-off energy, the particle trajectory is terminated. The choice of a tracking cut-off is determined by the residual range of particles. That is, the distance that particles with kinetic energy equal to the cut-off energy would travel in a given material. If, for example, a three-dimensional dose distribution is calculated with a spatial resolution of 2 mm, terminating particles with the residual range of 0.1 mm could be a good choice for the tracking cut-off. A similar concept is the so-called production threshold. In fact, numerical values of the tracking cut-offs and production thresholds often coincide. Their meaning, however, is different. In those interactions that produce secondary particles, only those particles are produced and tracked that have an energy exceeding the production threshold. Finally, for each particle type a list of interaction types should be created. And, for each interaction type a physics model has to be chosen, cross sections and all model parameters have to be determined for each material present in the simulation volume, and for all energies, from the tracking cut-off and production threshold, to the maximum energy.

Choosing Tallies

Another question that needs to be answered is as to what quantities are calculated and how. Solving problems of radiation transport with Monte Carlo simulations has two parts: generating particle trajectories and collecting information from particle tracks to estimate quantities of interest. The second part is called tallying. Tallies are discussed in detail in Sect. 4.3. At this point we will only mention that there are several types of tallies that can be used to estimate the same quantity, and a decision needs to be made as to which tallies to be used in the particular problem.

When calculating a three-dimensional dose distribution an important parameter is the spatial resolution of a dose grid. With a finer resolution achieving lower statistical uncertainties takes more computing time. On the other hand, with a coarse grid, some details of the dose distribution may be lost. For radiotherapy dose calculations a good resolution would be 1–4 mm, which is comparable to, or slightly more coarse than the spatial resolution of a computed tomography image of patient anatomy. An optimal grid would have a variable resolution, with a finer resolution where large dose gradients are expected and a coarse resolution elsewhere.

Other Considerations

The term variance reduction refers to a group of methods that reduce statistical uncertainties without introducing a bias, or systematic errors, to the solution. These methods are discussed in Sect. 4.4. A decision has to be made on what variance reduction methods will be used and how exactly. The latter includes choosing optimal parameters for the selected methods. A related question is calculating and reporting statistical uncertainties. It is important to have a plan that clarifies how statistical uncertainties will be calculated and reported. It should also be estimated in advance what level of statistical uncertainties can be achieved and what computing resources, especially the computing time, will be required. Another item to consider is what data will be included in the output files and how often will they be written to disk. Monte Carlo calculations often run for many hours and days. For this reason it is a common practice to periodically save intermediate results to disk. This saved data should include all information needed to restart the code and continue calculations in case they were terminated, accidentally or on purpose, before their full completion.

4.2 Neutral Particles

We start with neutral particles, because simulation of neutral particles transport is much simpler than that of charged particles. The entire next chapter is dedicated to simulation of charged particle transport.

4.2.1 Starting a Trajectory

The initial phase coordinates of the primary particle are sampled from the distribution given by the source function $S(\vec{r}, \vec{\Omega}, E)$. Often, but not always, $\vec{r}$, $\vec{\Omega}$, and E can be sampled independently from one another, that is, $S(\vec{r}, \vec{\Omega}, E) = S(\vec{r}\,)S(\vec{\Omega})S(E)$. If that is not the case, methods for sampling from joint distributions must be used. These methods were introduced in Sect. 2.12. Once the source emits a particle, the trajectory is generated until the particle is absorbed, leaves the computational domain, or its energy falls below the tracking cut-off. If secondary and higher generation particles are produced, they are handled in the same manner.

An important source type is the so-called phase space source. In the first calculation a scoring surface is defined and phase coordinates of all particles reaching that surface are saved in a "phase space file." In all subsequent calculations particle trajectories start at this scoring plane, with their initial phase coordinates read from the file. This method is intended for those problems where the first part of the system, closest to the source, is the same for all calculations and the rest of the system changes from one calculation to another. In that case, significant computing time can be saved, because there is no need to perform simulation in the first part of the system for every new configuration of the varying part of the system. A good example where this technique has been used is a simulation of radiation transport in a medical accelerator. For photon beam treatments, at a given nominal beam energy, for example 6 MV, the upper part of the accelerator is always the same. It is comprised of components, such as a bremsstrahlung target, a primary collimator, and a flattening filter that remain at fixed positions. The lower part of the accelerator beam line has movable parts, such as movable collimators (jaws) and a multileaf collimator, that define a beam aperture, and in some techniques, a fluence modulation pattern. Then, the surface defining the location of the phase space source can be a plane, normal to the beam, and located immediately below the flattening filter.

4.2.2 Stepping to the Next Collision Point

Homogeneous Medium

The distance l between two consecutive interaction points, or collisions, is called the free path. In a homogeneous medium it is exponentially distributed:

$$f_l(t) = \sigma \exp(-\sigma t), \tag{4.1}$$

where σ is the total cross section (cm^{-1}) and t is the distance (cm). An exponential distribution can be sampled using the inversion method (Sect. 2.2). The sampling formula is:

$$l = -\frac{1}{\sigma} \ln \gamma, \tag{4.2}$$

where γ is uniformly distributed between 0 and 1. Once the free path is sampled, the point of the next interaction $\vec{r}\,'$ can be found. If the particle started from point $\vec{r}$ in direction $\vec{\Omega}$, and the free path is l, then:

$$\vec{r}\,' = \vec{r} + \vec{\Omega} l, \tag{4.3}$$

or, in a componentwise form:

$$x' = x + \Omega_x l; \quad y' = y + \Omega_y l; \quad z' = z + \Omega_z l. \tag{4.4}$$

Next, a check needs to be performed as to whether or not point $\vec{r}\,'$ lies within the computational volume. If it does not, then the history of the particle is terminated, and the next particle starts. Otherwise, the particle is moved to point $\vec{r}\,'$, an interaction at this point is simulated, and, if the particle survives the interaction, then the next free path is sampled.

Heterogeneous Medium. General Theory

The probability of a free path exceeding distance t is (see Sect. 3.1):

$$P\{l > t\} = \exp\left[-\tau\left(\vec{r}, \vec{r}\,'\right)\right] \tag{4.5}$$

where

$$\tau\left(\vec{r}, \vec{r}\,'\right) = \int_0^t \sigma\left(\vec{r} + \vec{\Omega} t'\right) \mathrm{d}t'; \tag{4.6}$$

is the optical distance. Then, the cumulative distribution function (CDF) of the free path is

$$F(t) = 1 - \exp\left[-\tau\left(\vec{r}, \vec{r}\,'\right)\right]. \tag{4.7}$$

To sample a free path using the inversion method we first sample a γ and then find l by solving the following equation:

$$F(l) = \gamma. \tag{4.8}$$

Voxelized Heterogeneous Medium

If the medium is heterogeneous, the algorithm can be simplified by dividing the computational volume into small volumes, the so-called voxels, and assuming that within each voxel the medium is homogeneous.

Let us consider a particle crossing an interface between two voxels, each containing a different material, Fig. 4.1. The particle starts in material 1. First we sample its free path l_1 in material 1 assuming in this step that the material is infinite. If the sampled free path is less than the distance to the nearest voxel boundary in the direction of particle travel, $\vec{\Omega}$, this simply brings the particle to a new collision point that is inside voxel 1. If the free path exceeds the distance to the boundary then we find free path l_2 in material 2 by solving the equation

$$1 - \exp\left[-\tau\left(\vec{r}, \vec{r}\,'\right)\right] = 1 - \gamma. \tag{4.9}$$

Here, on the right-hand side, we wrote $1 - \gamma$ instead of γ to simplify slightly the algorithm. This does not introduce any error because γ and $1 - \gamma$ have exactly the same distribution. To calculate the optical distance τ we note that in the voxelized geometry σ is a piecewise constant function of $\vec{r}$. In this example, $\sigma(\vec{r}) = \sigma_1$ when the particle is in voxel 1 and $\sigma(\vec{r}) = \sigma_2$ when it is in voxel 2. After calculating the optical distance τ, and taking the logarithm of both sides of Eq. (4.9), we have:

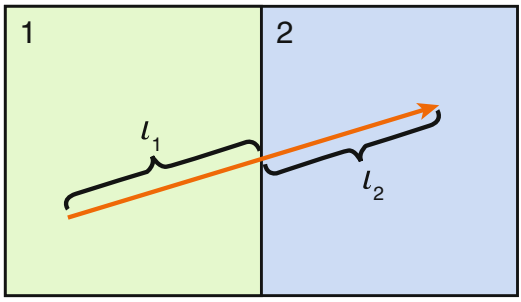

Fig. 4.1 Boundary crossing

$$-\sigma_1 l_1 - \sigma_2 l_2 = \ln \gamma, \tag{4.10}$$

where l_1 and l_2 are the distances the particle travels in voxels 1 and 2. Distance l_1 is already known, it is the distance from the starting point $\vec{r}$ to the boundary of voxel 1 in the direction $\vec{\Omega}$ the particle traveled. Distance l_2 is the free path in voxel 2. It is found by solving Eq. (4.10):

$$l_2 = -\frac{1}{\sigma_2} \ln \gamma - \frac{\sigma_1}{\sigma_2} l_1. \tag{4.11}$$

Next, a check must be performed as to whether or not the particle traveling distance l_2 reaches the next boundary, for example, between voxels 2 and 3. If it does, $(-\sigma_3 l_3)$ is added to the left-hand side of Eq. (4.10) and l_2 is replaced by the distance the particles traveled within voxel 2. Then, this modified equation is solved for l_3. If the particle reaches yet another boundary, this procedure is repeated.

Maximum Cross-Section Method

This method works best when for any given energy, differences in cross sections between different materials are small. This method does not require a voxelized geometry.

Algorithm Groundwork

Find the maximum cross section $\sigma_{\max}(E)$, i.e., a cross section that for any point $\vec{r}$ in the computational domain satisfies:

$$\sigma_{\max}(E) \geq \sigma\left(\vec{r}, E\right). \tag{4.12}$$

Note, that $\sigma_{\max}$ is a function of energy E, but it does not depend on $\vec{r}$. For brevity we leave energy E out.

Algorithm Overview

The free path l is sampled from an exponential distribution

$$f_l(t) = \sigma_{\max} \exp(-\sigma_{\max} t). \tag{4.13}$$

This step is no different from sampling the free path in a homogeneous medium. On average, the free path sampled from this distribution is shorter than the real one, and the number of interactions is higher. This bias, however, is compensated exactly by performing the following. Once the particle arrives at the new interaction point,

a random number is drawn to determine whether this interaction is real or it is a δ-scattering. By definition, in δ-scattering neither particle energy nor its direction changes. This is a fictitious interaction that is equivalent to no interaction occurring at all. The probability of δ-scattering is given below (step 5 of the algorithm). To summarize, there will be more interactions than in reality, but there will be no bias, because only a fraction will be real interactions, and the rest will be δ-scattering.

Algorithm

1. Sample free path l from distribution $\sigma_{\max} \exp(-\sigma_{\max} l)$.
2. Move the particle to the next point of interaction at $\vec{r}\,' = \vec{r} + \vec{\Omega}\, l$.
3. If $\vec{r}\,'$ is outside the computational domain, terminate the particle history.
4. Otherwise, determine what material is at point $\vec{r}\,'$, find the cross section for this material, $\sigma(\vec{r}\,')$.
5. Calculate the probability of δ-scattering, $P_\delta = \left[\sigma_{\max} - \sigma(\vec{r}\,')\right]/\sigma_{\max}$.
6. Sample the type of interaction, δ-scattering or real. Generate a γ. If $\gamma < P_\delta$, then it is a δ-scattering, otherwise it is a real interaction.
7. If it is a δ-scattering, nothing happens, go to step 1, and sample the next free path.
8. If it is a real interaction, simulate it. The particle momentum may change, or it may be absorbed, and secondary particles may be produced. Proceed depending on the outcome of the interaction.

Proof

We now prove that the above algorithm produces a distribution of path length between real collisions that satisfies Eq. (4.5). The probability of a particle traveling a distance $t + \mathrm{d}t$ without interactions is the product of the probability of traveling a distance t without interactions and the probability of zero interactions within $\mathrm{d}t$:

$$P(l > t + \mathrm{d}t) = P(l > t)\, P(n = 0|\mathrm{d}t), \tag{4.14}$$

where n is the number or real interactions. In the algorithm, the free path is sampled using the maximum cross section and at the collision point either a real interaction or a δ-scattering takes place. As the particle travels a distance $\mathrm{d}t$ the probability of a collision, which may or may not be real, is $\sigma_{\max}\mathrm{d}t$. Accordingly, the probability of no collisions is $1 - \sigma_{\max}\mathrm{d}t$. If a collision does occur within $\mathrm{d}t$, the probability that it is not a real one is P_δ. If no collisions occur then, obviously, the probability of a real interaction is zero. This brings us to a total probability equation given by:

$$P(n = 0|\mathrm{d}t) = \sigma_{\max}\mathrm{d}t \cdot P_\delta + (1 - \sigma_{\max}\mathrm{d}t) \cdot 1, \tag{4.15}$$

where

$$P_\delta = \frac{\sigma_{\max} - \sigma(\vec{r}\,')}{\sigma_{\max}}; \quad \vec{r}\,' = \vec{r} + \vec{\Omega}t, \tag{4.16}$$

as we defined earlier. We now substitute Eqs. (4.15) and (4.16) in Eq. (4.14) to arrive at a simple differential equation for $P(l > t)$:

$$\frac{\mathrm{d}}{\mathrm{d}t} P(l > t) = -\sigma\left(\vec{r}\,'\right) P(l > t)\,. \tag{4.17}$$

Solving this equation with the boundary condition $P(l > 0) = 1$ produces Eq. (4.5). □

4.2.3 Interaction

After the free path is sampled and the particle has moved to a new interaction point, the type of interaction is sampled. The type of interaction is a discrete random number. If the total number of interaction types is k, then the distribution of this random number is given by probabilities: $p_1, p_2, \ldots, p_k$. The probability p_i, is the ratio of the cross section for interaction of type i to the total cross section:

$$p_i = \frac{\sigma_i}{\sigma}; \quad i = 1, 2, \ldots, k; \tag{4.18}$$

$$\sigma = \sigma_1 + \sigma_2 + \cdots + \sigma_k. \tag{4.19}$$

This brings us to the standard problem of sampling a discrete random number. We have addressed it in Sect. 2.3.

Example

Gamma radiation (photons). The three main types of interaction are:

1. Photoelectric absorption. The photon disappears, a photoelectron is produced, and, as a result of atomic de-excitation, Auger electrons or fluorescent photons are emitted.
2. Compton scattering. The photon is scattered and a Compton electron is produced.
3. Pair production. It occurs only when photon energy exceeds $2m_e c^2 = 1.022$ MeV. The photon disappears, an electron and a positron are produced. The positron soon annihilates producing two photons.

From the simulation viewpoint, interactions can be classified into the following four types based on what happens to the primary particle:

1. Absorption. Particle history is terminated.
2. Scattering. Particle momentum changes. If particle energy falls below the tracking cut-off, its history is terminated. Otherwise, the particle, now with a new momentum, travels to the next interaction point.
3. Conversion. The primary particle disappears and a new particle, or particles, of a different type, are produced.
4. Multiplication. After the collision, two or more particles of the same type as the primary particle emerge. According to the quantum identity principle it is impossible to determine which of the emerged particles, if any, is the primary particle. For simulation purposes a common convention is that the most energetic particle is the primary particle.

When a particle history is terminated, it is recommended to run an end-of-history procedure. For example, in dose calculations it can process the residual energy of the particle so as to ensure energy conservation. Any of the above interaction types can in principle produce secondary particles. Moreover, after a particle interacts with an atom, its electron shell or nucleus may undergo a transition to an excited state. De-excitation may produce energetic particles, for example photons and electrons, capable of ionizing the medium. If secondary particles are produced, their parameters are stored in computer memory. Usually a stack-type array is used for that purpose. The list of stored quantities should normally include: particle type, weight, x, y, and z of its origin, and initial energy and direction, $\vec{\Omega}$. Here, the weight does not refer to the particle mass. It is an additional particle parameter used in variance reduction techniques. Also, after a particle history is terminated, an initial phase coordinate of the next particle is generated. To this end, the stack is checked for particles that have not been simulated yet. If such particles are found, the initial phase coordinate is read from the stack. If more than one particles are present, parameters of the particle with the minimum energy should be read first. This strategy is called the lexicographic scheme. Only after histories of all particles from the stack have been simulated, a new particle can be emitted by the source.

4.3 Tallies

So far in this chapter, we have been mostly concerned with generating particle trajectories. However, trajectories themselves are rarely the purpose of calculations. As we mentioned in Chap. 3, our purpose, in most general terms is the calculation of a detector reading J, for example dose or fluence at a given point, or more often, their spatial distributions. The process of extracting information from generated particle tracks to estimate an observable J is called tallying, and the estimators are called tallies. In this section we introduce several basic tallies.

4.3.1 Surface Crossing Tally

In Sect. 3.1, when we introduced Definition 1 of fluence, we wrote

$$\mathrm{d}N\left(\vec{r},\vec{\Omega},E\right)=\Phi\left(\vec{r},\vec{\Omega},E\right)\left|\vec{\Omega}\cdot\vec{n}\right|\mathrm{d}A. \tag{4.20}$$

The meaning of this equation is illustrated in Fig. (3.3). If we solve this equation for fluence

$$\Phi\left(\vec{r},\vec{\Omega},E\right)=\frac{1}{|\vec{\Omega}\cdot\vec{n}|}\frac{\mathrm{d}N}{\mathrm{d}A};\qquad|\vec{\Omega}\cdot\vec{n}|=|\cos\theta|=|\mu|, \tag{4.21}$$

we can see that to estimate fluence of particles traveling in direction $\vec{\Omega}$, we need to calculate the average number of particles per unit area of the surface and divide the result by the absolute value of the cosine of the angle between the direction the particle travels $\vec{\Omega}$ and normal $\vec{n}$ to the surface.

Then, to calculate fluence at a point $\vec{r}$, we place at this point a small flat surface ΔA, generate particle trajectories, and calculate fluence using the following equation:

$$\Phi=\frac{1}{\Delta A}\sum_{i=1}^{\nu}\frac{1}{|\mu_i|}, \tag{4.22}$$

where the summation is over all particles that crossed the surface. If we count all particles that crossed the surface, then the result is an estimate of the total fluence. If we count only particles traveling in a certain direction, or with a certain energy, then the result is the angular or energy distribution of fluence, respectively. Practically, however, we can only count particles whose vector $\vec{\Omega}$ is within a finite solid angle $\Delta\omega$ and whose energy E is within a finite energy interval ΔE. Equation (4.22) defines the surface crossing tally for fluence. To tally any other observable J, the 1 in the numerator should be replaced by the corresponding detector response function $D(x_i)$.

4.3.2 Boundary Crossing Tally

Surface tallies estimate fluence or other quantities on a surface. The boundary crossing tally, on the other hand, is intended for calculating quantities averaged over a finite, usually small, volume ΔV. The word boundary here refers to the boundary Γ of the volume ΔV. As usual, the tally is a sum. Contributions to the sum are calculated when a particle crosses the boundary, either entering or leaving the volume ΔV of interest.

Boundary crossing tally is particularly useful for dose, or energy deposition calculations. If a particle enters the volume ΔV with energy E and leaves it with energy E', then the energy Q deposited in ΔV is $E - E'$. If the particle does not exit the volume, then $Q = E$. If one particle enters, but several exit, then we need to subtract the total energy of all exiting particles, $Q = E - \sum_{i=1}^{\nu} E'_i$. Dose is the average deposited energy Q divided by the mass of material in ΔV. Note, that this is not the dose at a point, but the dose averaged over ΔV. Obviously, to calculate the average Q, we need to generate trajectories of many particles, calculate energy deposited by each particle, and then take the average of all the deposits. The latter includes zero deposits for those particles that never reached the volume, or traversed it without inelastic interactions. The described procedure is the boundary crossing tally for dose or energy deposited. When a spatial dose distribution is calculated, usually, the entire computational domain, or a large portion of it, is divided into a large number of small volumes called voxels, and for each voxel a separate dose tally is calculated.

4.3.3 Collision Tally

Let us consider a small volume ΔV. The average number of collisions in it, $N(\Delta V)$, is proportional to the collision density $\sigma\Phi$:

$$N(\Delta V) = \sigma\Phi\Delta V. \tag{4.23}$$

Then, the fluence averaged over ΔV can be estimated using the equation

$$\Phi = \frac{1}{\Delta V}\sum_{i=1}^{\nu}\frac{1}{\sigma_i}, \tag{4.24}$$

where the summation is over all collisions in ΔV, and σ_i is the total cross section of a particle undergoing i-th collision. The cross section is calculated for particle energy immediately before the collision. This formula defines the collision tally for fluence. Fluence on the left-hand side of the above equation could be the total fluence, or fluence differential in angle or energy, depending on what filters are applied when collisions are counted.

Energy Q deposited in a volume ΔV can also be calculated as a sum over collisions. The tally is obvious and given by:

$$Q(\Delta V) = \sum_{i=1}^{\nu}\epsilon_i, \tag{4.25}$$

where ϵ_i is the energy deposited in i-th collision. It is calculated as the difference between the particle energy before the collision and after collision, and the total

energy of all particles produced in the collision. In ionization collisions, for example, ϵ_i can be approximated by the ionization potential of the electron shell involved in the interaction. This approximation, however, is not applicable when ionization is followed by atomic de-excitation through fluorescence, emission of Auger electrons, or other similar channels.

4.3.4 Track End Tally

This tally is invoked when particle trajectories are terminated. Its applications are limited. We will give two examples.

Dose Calculation

If particle energy falls below the tracking cut-off, the trajectory is terminated. At this point, the particle still has a certain amount of energy that must be included in dose tally. In boundary crossing tally this residual energy is accounted for automatically. For other tallies, this is not always the case, depending on tally type and its implementation. Track end tally must be calculated, if no other tally accounts for the residual energy. The same considerations also apply to particle absorption.

Ion Implantation Profiles

Ion implantation is an important technology used in semiconductor device fabrication. Spatial distribution of concentration of implanted ions largely determines properties of a semiconductor. This distribution is conveniently characterized by depth and lateral ion implantation profiles. If these profiles are calculated with Monte Carlo simulations, then the track end tally must be used. This is because in this case we need to register the location, usually a voxel, where the ion stops.

4.3.5 Path Length Tally

According to Definition 2 of fluence (Sect. 3.1), fluence $\Phi(x)$ is the total distance per unit volume traveled by particles with phase coordinate x. Directly from this definition, it follows yet another method for estimating fluence, in which it is averaged over a small volume ΔV. This tally, called the path length tally, is defined by the following equation:

$$\Phi = \frac{1}{\Delta V} \sum_{i=1}^{v} l_i \ , \tag{4.26}$$

where the summation is over all free paths of all particles that entered ΔV, and l_i is the part of a free path that belongs to ΔV.

Example: Electron Slowing Down Spectrum

The slowing down spectrum $\Phi(E)$ is a solution of the energy degradation equation that we introduced in Sect. 3.9.3, Eq. (3.112). The Monte Carlo algorithm for solving the equation is based on the use of path length tally. In this case, the fluence depends only on electron energy and is spatially uniform. This means that we do not need to track particle coordinates, or its direction of travel, and hence we can leave out the volume ΔV.

Algorithm Groundwork

Determine the minimum and maximum possible electron energies, $E_{\min}$ and $E_{\max}$. Normally, $E_{\min}$ is equal to the tracking cut-off, and $E_{\max}$ is the maximum energy of the electron source. Divide the entire energy range, from $E_{\min}$ to $E_{\max}$, into energy bins $\Delta E_i, i = 1, 2, \ldots, k$. Boundaries of the energy bins should be equally spaced on a logarithmic scale. Initialize the histogram array by setting $h_i = 0, i = 1, 2, \ldots, k$.

Algorithm

It is written for an event-by-event technique, because these calculations are relatively fast, hence there is no need to use condensed history methods. The following are the calculations performed at one step of an electron trajectory.

1. Sample electron energy E' immediately after an interaction. If this is the first step, sample E' from the energy distribution of the source. If E' falls below the tracking cut-off, terminate the history.
2. Sample free path l using the total cross section at energy E', $\sigma(E')$.
3. Find the energy bin ΔE_j to which energy E' belongs, $E' \in \Delta E_j$.
4. Add the sampled free path length l to the corresponding histogram bin: $h_j = h_j + l$.
5. Simulate an interaction, make the next step.

After a sufficiently large number of histories have been generated, simulations are terminated, and the histogram estimate, Φ_i, of the slowing down spectrum is calculated

$$\Phi_i = \frac{h_i}{N\Delta E_i}, \quad i = 1, 2, \ldots, k, \tag{4.27}$$

where N is the number of histories. The units for the calculated fluence spectrum are length over energy, e.g., nm/eV. It is normalized per source particle, and its integral over energy is equal to the total path length of the primary electron and all the secondary and higher generation electrons it produced. The number of electrons produced by the primary electron is very large, and they tend to have low energies. For this reason, statistical uncertainties at the low energy end of the slowing down spectrum are much lower than those at the high energy end. This property can be used to speed up the calculations. For example, first a relatively small number of histories can be run with the minimal cut off energy, until a good accuracy for the low energy part of the spectrum is achieved. Then, the cut-off energy can be increased, and a larger number of histories are run to obtain the rest of the spectrum. Of course, with a higher cut off energy, computing time per history will be much lower. Finally, the results of these two calculations can be combined with appropriate weights. Alternatively, a floating cut off can be used. For example, after each batch of histories, the uncertainties are evaluated and, depending on the result, the cut off is moved up or down on the energy scale. To convert the histogram into fluence spectrum correctly, this approach will require counting and storing the number of histories, separately for each energy bin.

4.3.6 Adjoint Function Tally

The adjoint function $\Phi^+(x)$ was defined in Sect. 3.1 as the average contribution to an observable J from a particle originating at a phase point x. If, for example, fluence is the observable, then $\Phi^+(x)$ is the fluence from a particle originated at x. Let us consider the following problem.

Problem

For a source $S(x)$, calculate the fluence $\Phi_{\Delta A}$ across a planar surface ΔA. The medium is homogeneous, and the total cross section is $\sigma(E)$.

We have already introduced several tally types for this type of problems. Here we introduce one more. It uses the unscattered adjoint function, $\Phi_0^+(x)$, that often can be found analytically. Figure 4.2 illustrates how this new tally is used. From the definition of Φ^+, we have:

$$\Phi_{\Delta A} = \int \mathrm{d}x_0 S(x_0)\, \Phi^+(x_0)\,, \tag{4.28}$$

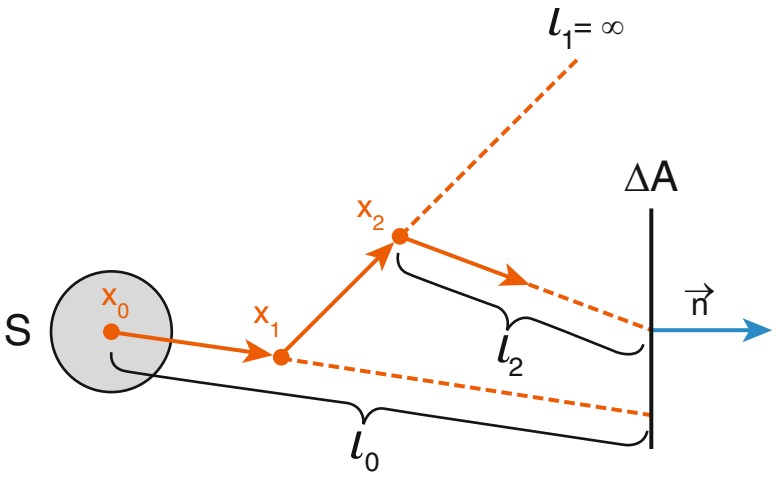

Fig. 4.2 Adjoint function tally

where $\Phi^+(x_0)$ is the contribution to the fluence $\Phi_{\Delta A}$ from particles that started with a phase coordinate x_0. Then, to estimate the integral in Eq. (4.28) we need to sample x_0 from the distribution $S(x_0)$, and then calculate $\Phi^+(x_0)$. For the latter calculation, we could use the Monte Carlo method, and for example the surface tally. But there is another way of using the Monte Carlo method for this calculation. At the first step of the trajectory, we apply the total probability equation

$$\Phi^+(x_0) = P_0(x_0)\,\Phi_0^+(x_0) + [1 - P_0(x_0)]\,\Phi^+(x_0|\text{collision})\,, \tag{4.29}$$

where $P_0(x_0)$ is the probability that a particle starting from x_0 reaches the surface ΔA unscattered, and $\Phi^+(x_0|\text{collision})$ is the adjoint function for a particle that also started at point x_0, but interacted at least once before reaching surface ΔA. The first term on the right-hand side can be calculated analytically, yielding

$$P_0(x_0)\,\Phi_0^+(x_0) = \exp\left[-\sigma(E_0)\,l_0\right]\frac{1}{\Delta A|\mu_0|};\quad \mu_0 = \left(\vec{\Omega}_0 \cdot \vec{n}\right), \tag{4.30}$$

where $\vec{n}$ is the normal to the surface ΔA. The result does not depend on which of the two normal directions was chosen, because here the absolute value of the cosine is taken. To calculate the second term on the right-hand side of Eq. (4.29), we move the particle to the first interaction point, sample its phase coordinate, x_1, after the interaction, and then repeat exactly the same steps as for point x_0. This will add another analytical term, this time $P_0(x_1)\Phi_0^+(x_1)$. We continue these steps until the last interaction point, the point number ν. This procedure produces the following estimate for fluence (see Fig. 4.2):

$$\Phi_{\Delta A} = \sum_{i=0}^{\nu-1} P_0\left(x_i\right) \Phi_0^+\left(x_i\right) = \sum_{i=0}^{\nu-1} \exp\left[-\sigma\left(E_i\right) l_i\right] \frac{1}{\Delta A |\mu_i|}. \tag{4.31}$$

This equation defines the adjoint function tally for fluence. Note that the last point of the trajectory, point ν, does not contribute to the sum, because it is the last point, after reaching it the particle will not contribute to fluence.

Adjoint function tally is similar to collision tally in that it is a sum over particle collisions. There are, however, significant differences. First, contributions to the tally from a collision are different. Second, the adjoint function tally uses particle phase coordinate after the collision, whereas collision tally uses particle energy before the collision. Finally, adjoint function tally includes a contribution from x_0 (at the source) and excludes that from x_ν (the last point). It is vice versa for collision tally.

4.3.7 Other Tallies

In this section, we overview a few other, frequently used, but more specialized tallies.

Voxel Tally

In this technique, the computational domain is divided into small cuboids, called voxels. For each voxel a material must be defined. In a heterogeneous medium, there could be more than one material in a voxel, and the density in a voxel could be nonuniform. If the voxel is sufficiently small, variations in material properties within a voxel can be neglected, and a uniform material can be assigned to each voxel. This involves averaging material properties over the voxel volume, and introduces systematic errors.

This technique is often used for calculations of patient dose distributions in external beam radiotherapy. In this case, the anatomical information is presented in the form of a computed tomography (CT) image of the patient. Other imaging modalities can also be used in the treatment planning process. They usually provide complementary information, helping in locating the tumor, or measuring the extent of its motion. To perform Monte Carlo calculations, the CT image has to be converted into a voxelized phantom. An important step in the conversion procedure is translating the image data into material properties. More specifically, in the CT image anatomical information is given in the Hounsfield units (HU). For Monte Carlo calculations an algorithm that for any given HU number outputs the material chemical composition and its density is required. In this technique, voxels not only define the geometry and materials, but also are usually used to tally the dose. Then, choosing the voxel size is an optimization problem of finding a balance

between statistical uncertainties and systematic errors caused by voxelization of the geometry. For patient dose calculations the voxel size is typically between 2 and 4 mm, and is uniform throughout the phantom. This type of calculation should be performed for research purposes only, and not for planning or guiding any actual treatments.

Before a Monte Carlo system can be used for patient dose calculations, at least some basic tests must be performed. They should include calculation of dose distributions in a water phantom and comparing the results with measurements. In external beam radiotherapy, the dose distributions are characterized by depth dose curves and lateral profiles. The depth dose is simply dose as a function of depth. It is normally measured on the central axis of the beam and for a range of field sizes. The lateral dose profile is the dose as function of distance from the central axis of the beam, measured in a plane normal to the central axis. Dose profiles are measured at several depths and for a range of field sizes.

Examples of measured depth dose curves and lateral profiles are shown in Figs. 4.3 and 4.4, respectively. An important property of depth dose curves is a steep dose gradient at the depths of up to 1–3 cm depending on beam energy. Dose calculations in this, so-called build-up region, require a voxel size of 2 mm or less in the depth direction. The voxel size can be gradually increased beyond this region, but it should not exceed 4–5 mm. The optimal voxel size in the lateral direction depends on dose variation in that direction. The latter can be seen in the lateral dose profiles. In a conventional accelerator, a flattening filter is placed in the beam line to achieve a more uniform fluence profile. In this case, the voxel size in the lateral direction can be as large as 1×1 cm, except near the beam edges, where dose gradient is high. In more recent accelerator models, the flattening filter can be removed to produce the so-called flattening filter free (FFF) beams. The effect of removing the flattening filter on lateral profiles can be seen in Fig. 4.4. For FFF beams, the lateral size of the voxels should be reduced. In this case the degree of

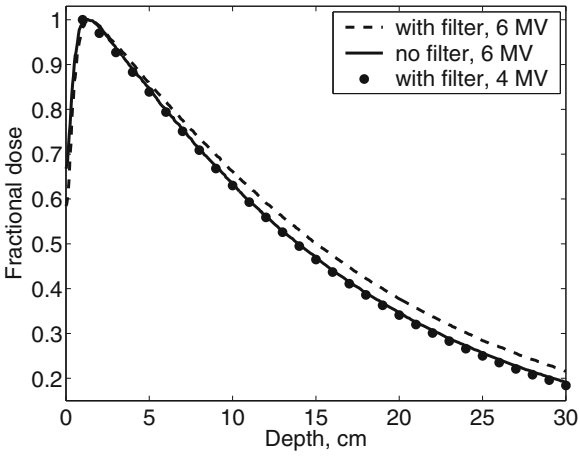

Fig. 4.3 Measured depth dose dependence for 10×10 cm photon beams (Vassiliev et al. 2006): 6 MV with a flattening filter (*dashed line*), 6 MV FFF (*solid line*), 4 MV with a filter (*circles*).

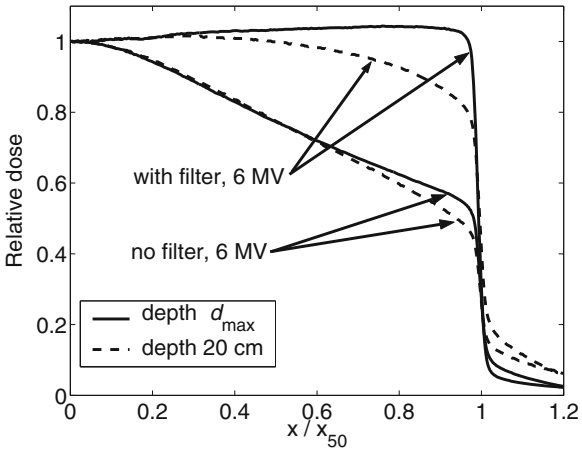

Fig. 4.4 Measured lateral dose profiles (Vassiliev et al. 2006): 6 MV photon beams with (*two upper curves*) and without (*two lower curves*) a flattening filter. The field size is 40 × 40 cm. The depths are 1.4 cm (*no filter, solid line*), 1.5 cm (*with filter, solid line*), and 20 cm (*with and without filter, dashed lines*); x is the distance from the central axis, x_{50} is the distance where the dose (*with the filter*) drops to 50 % of dose on the central axis.

beam non-flatness increases with increasing beam energy, and voxels sizes should be adjusted accordingly. In lateral dose profiles, there is a very steep dose gradient in the penumbra area, i.e., in the area around the point x_{50}, where the dose drops to 50 % of the dose on the central axis. In the penumbra region, the voxel size in the lateral direction should not be larger than 1–2 mm. It can be gradually increased away from this region, perhaps up to about 5 mm provided that the profile is sufficiently flat. It should be kept in mind that the location of the penumbra changes with depth, because the beams are divergent. The voxel size in the depth direction for lateral profiles should not exceed 2–4 mm.

Mesh Tallies

In the voxel tally method the material geometry and tallies are defined on the same spatial grid. This is not always an optimal configuration. For example, tallies may require a higher spatial resolution grid than the grid needed for defining materials. In the mesh tally method, the material geometry and tallies are defined on two generally different grids. Neither grid has to be rectangular. A mesh for tallying dose can be, for example, cylindrical or spherical. A mesh tally can define either surface or volume tally types. A downside of using mesh tallies for dose calculations is that, in this case, one needs to know the mass of matter in each volume defined by the mesh tally, but because materials are defined on a different spatial grid, a volume defined by the mesh tally may contain several different materials.

Current Tally

The current j is sometimes also referred to as the "planar fluence." It does not appear anywhere in the formalism presented in this book. It, however, has been mentioned in the literature (Lewis and Miller 1993), and the tally has been implemented in some standard Monte Carlo software. The current tally is similar to the surface tally for fluence, Eq. (4.22), except that it does not include the cosine:

$$j = \frac{1}{\Delta A} \sum_{i=1}^{N} 1. \tag{4.32}$$

Here the sum is over all particles crossing the surface ΔA. The tally gives the net number of particles crossing the surface, per unit area. It is highly unlikely that this tally will be needed for radiotherapy applications. It certainly cannot be used as a substitute for fluence tally. One example where the current tally may by appropriate is in problems involving charged particles, electrons for example, and the quantity that needs to be calculated is the electric current across a surface. Current can also be defined as a vector $\vec{j}$ that is parallel to $\vec{\Omega}$. It can also be a signed scalar, i.e., negative or positive depending on the direction from which the particle reaches the surface, or the particle charge.

Pulse Height Tally

It is similar to the boundary crossing estimator of deposited energy. The difference is that the deposited energy estimator outputs the average energy deposited in a volume ΔV, whereas pulse height tally outputs a histogram. The entire range of possible energy deposits is divided into energy bins. When the energy deposited by an individual particle (referred to as "pulse height") is determined, and it falls within an energy bin j, the number of counts in bin j is incremented by 1. Variance reduction techniques (Sect. 4.4) generally cannot be used for this tally type, because the Boltzmann equation is not applicable to fluctuational characteristics, such as pulse height distributions.

4.4 Variance Reduction

4.4.1 Algorithm Efficiency

The general Monte Carlo process involves the following steps:

1. Generate a large number, N, of particle histories.
2. Each history produces an estimate ξ_i of a quantity of interest, such as the dose in a voxel.

3. Estimate the expectation of ξ_i using the sample average

$$\mathrm{E}\{\xi\} = \frac{1}{N}\sum_{i=1}^{N}\xi_i. \tag{4.33}$$

   Because $N \gg 1$, the sample average is usually normally distributed, according to the central limit theorem. There are, however, exceptions where the theorem is not applicable, for example when the variance of ξ_i is infinite.
4. Estimate the variance of the sample average:

$$\mathrm{Var}\left\{\frac{1}{N}\sum_{i=1}^{N}\xi_i\right\} = \frac{\mathrm{Var}\{\xi_i\}}{N}. \tag{4.34}$$

5. Choose a confidence level p, find a confidence interval. For example, if $p = 0.95$, then the confidence intervals is (Appendix A)

$$\left[\frac{1}{N}\sum_{i=1}^{N}\xi_i - \frac{1.96\sqrt{\mathrm{Var}\{\xi_i\}}}{\sqrt{N}};\ \frac{1}{N}\sum_{i=1}^{N}\xi_i + \frac{1.96\sqrt{\mathrm{Var}\{\xi_i\}}}{\sqrt{N}}\right]. \tag{4.35}$$

   In other words, at the confidence level of 0.95 the uncertainty is approximately 2 standard deviations.

The above scheme is, of course, simplified, and deviations from it are possible. From the above equation, the uncertainty is

$$\delta = \frac{1.96\sqrt{\mathrm{Var}\{\xi_i\}}}{\sqrt{N}}. \tag{4.36}$$

The sample size or the number of histories N_δ needed to achieve a given level of uncertainties is then

$$N_\delta = \frac{1.96^2\mathrm{Var}\{\xi_i\}}{\delta^2}. \tag{4.37}$$

Accordingly, the total CPU time T needed to achieve an uncertainty level δ is

$$T = tN_\delta = \frac{t\cdot\mathrm{Var}\{\xi_i\}}{\delta^2}1.96^2, \tag{4.38}$$

where t is the average CPU time needed to generate one ξ_i, i.e., time per history.

Then, an efficient algorithm will have a small value of the product $t\cdot\mathrm{Var}\{\xi\}$, and the efficiency ϵ of a Monte Carlo algorithm can be defined as

$$\epsilon = \left(t\cdot\mathrm{Var}\{\xi_i\}\right)^{-1}. \tag{4.39}$$

Normally, $\mathrm{Var}\{\xi_i\}$ is not known, but it can be replaced by the sample variance s^2. Ideally, we would like to reduce both t and $\mathrm{Var}\{\xi_i\}$. In reality, the focus is on achieving a significant reduction in $\mathrm{Var}\{\xi_i\}$ that does not slow down calculations too much, so that the product $t \cdot \mathrm{Var}\{\xi_i\}$ is reduced.

If, after an algorithm modification, the above goal is achieved, **and the expectation** $\mathrm{E}\{\xi\}$ **remains exactly the same**, i.e., no systematic errors are introduced, then the technique is referred to as variance reduction.

4.4.2 Particle Splitting

There are two similar methods of particle splitting. In method 1, when a particle undergoes an interaction that produces one or more secondary particles, the number of secondary particles is artificially increased by a factor of n, as compared with the real process. All the secondary particles are sampled from the same distribution, and each new particle has a different (random) momentum. To compensate for the bias in the number of secondary particles, each particle is assigned a weight of $1/n$. For example, in a medical accelerator, an electron is incident on a tungsten target and produces bremsstrahlung photons, as shown in Fig. 4.5. These photons deliver the treatment dose. If in a collision the electron produces one photon, in a Monte Carlo algorithm we can, instead, produce for example five, each with a weight of 1/5. The benefit is that there are more photons produced per incident electron. This saves the CPU time, because simulation of electron trajectories is relatively slow, but the dose to a patient is delivered mostly by photons.

In a variation of this technique, all five bremsstrahlung photons would have exactly the same momentum. This saves some time, because instead of five times, the photon momentum is sampled once. The downside is that the trajectories of the five photons are correlated, especially near the point of origin.

In method 2, a particle is split into $n > 1$ identical particles as it enters a volume. All daughter particles originate at the point of entry on the volume boundary and have the same momentum as the parent particle. To compensate for the bias, each of the n particles is assigned a weight of $1/n$.

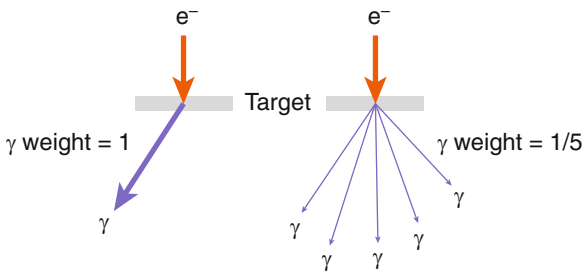

Fig. 4.5 Particle splitting

The purpose of particle splitting is to increase the number of particles that are more likely to contribute to detector reading, and not of those particles that have little chance of reaching it. Particle splitting does not necessarily improve the algorithm efficiency, because of the additional CPU time needed to generate trajectories of extra particles. For this reason particle splitting should be limited to a relatively small volume, perhaps around the detector. It would also help, if trajectories of extra particles were terminated when they travel far enough from the detector. However, simply terminating particle histories at a certain distance from the detector would introduce a systematic error. The Russian roulette technique presented in the next subsection achieves the same result, i.e., termination of particle histories where it is needed, but with the bias compensated exactly by particle weights.

4.4.3 Russian Roulette

A Russian roulette technique is, in a sense, the opposite of particle splitting. Its purpose is to kill particles that are unlikely to contribute to detector reading, which obviously saves computing time. Similarly to particle splitting, Russian roulette can be played at a collision point, when secondary particles are produced, or when a particles enters a volume. In either case, a survival probability $p_s < 1$ must be specified for each particle. Then a random number γ uniformly distributed in $(0, 1)$ is generated, and if $\gamma > p_s$, then the particle history is terminated. Otherwise, the particle survives and its weight increases by the factor of $1/p_s$.

The particle splitting technique and Russian roulette can be used to limit fluctuations of particle weights when particle trajectories are biased. In a technique called weight window, weights are restricted to a user defined range ("window") from $w_{\min}$ to $w_{\max}$. Particles with weights lower than $w_{\min}$ are eliminated using Russian roulette. In this case a low weight particle is either killed or, if it survives, its weight increases. Similarly, particles with weights exceeding $w_{\max}$ are split, which results in weight reduction.

4.4.4 Forced Collisions

If we need to increase the average number of interactions in a certain volume V, we can artificially increase interaction cross sections in this volume, and compensate for the bias by weights Q_n defined in Sect. 3.10. The forced interactions technique offers an alternative approach. In this approach a particle entering the volume interacts within it at least once with the probability p_c (collision probability) defined by the user.

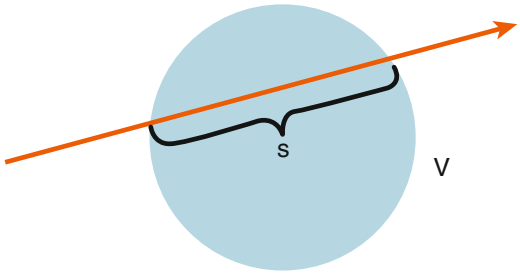

Fig. 4.6 Forced interactions

Let us consider a simple case of a homogeneous material. The free path length is distributed exponentially between zero and infinity:

$$p\,(l) = \sigma \exp\,(-\sigma l);\;\; l \geq 0. \tag{4.40}$$

We then introduce a biased distribution $\tilde{p}(l)$, where the exponential distribution is modified so that the particle interacts within the volume with probability p_c. We will assume that when $l = 0$ the particle is on the surface of the volume, entering it, and that s is the projected chord length of the particle within volume V, as shown in Fig. 4.6. The biased distribution is

$$\tilde{p}\,(l) = p_c f_1\,(l) + (1 - p_c) f_2\,(l)\,;\;\; 0 < p_c < 1, \tag{4.41}$$

where

$$f_1\,(l) = \begin{cases} A_1 \exp\,(-\sigma l), & 0 \leq l < s; \\ 0, & l \geq s \;\text{ and }\; l < 0; \end{cases} \tag{4.42}$$

and

$$f_2\,(l) = \begin{cases} A_2 \exp\,(-\sigma l), & l \geq s; \\ 0, & l < s. \end{cases} \tag{4.43}$$

To sample from the biased distribution $\tilde{p}(l)$ we can use the algorithm for sampling a sum of distributions that we discussed in Sect. 2.7. According to this algorithm, we sample with probability p_c from distribution $f_1(l)$ and with probability $1 - p_c$ from distribution $f_2(l)$. Distribution $f_1(l)$ is defined so that it is zero everywhere except within volume V. Hence, with probability p_c the interaction occurs within V. Similarly, with probability $1 - p_c$ the interaction point occurs beyond point $l = s$, that is outside V.

The normalization constants A_1 and A_2 can be found from

$$\int_0^\infty \mathrm{d}l\, f_1\,(l) = \int_0^\infty \mathrm{d}l\, f_2\,(l) = 1, \tag{4.44}$$

resulting in

$$A_1 = \frac{\sigma}{1 - \exp(-\sigma s)}; \quad A_2 = \sigma \exp(\sigma s). \tag{4.45}$$

Because we are biasing the particle trajectory, specifically the distribution of free path, we need to calculate the weight and assign it to the particle. For each particle the weight is calculated only once, because only one step of the particle trajectory is biased, that is the step that begins on the volume surface, when the particle enters the volume. According to Eq. (3.175) that defines the weight, the weight is equal to the ratio of the kernel $k_F(x' \rightarrow x)$ to the transition probability of the Markov process $p(x' \rightarrow x)$. Our transition probability coincides with the kernel, except that the exponent of the form given by Eq. (4.40) is now replaced by $\tilde{p}(l)$ given by Eqs. (4.41)–(4.43). Hence, the particle weight is

$$w(l) = \frac{p(l)}{\tilde{p}(l)} = \begin{cases} [1 - \exp(-\sigma s)]/p_c, & l < s; \\ \exp(-\sigma s)/(1 - p_c), & l \geq s. \end{cases} \tag{4.46}$$

The weight is not defined for $p_c = 1$ and $p_c = 0$. This brings us to the following simple algorithm.

Algorithm

1. Generate a γ (i.e., a random number uniformly distributed in (0,1)).
2. If $\gamma < p_c$, then sample free path l from the distribution $f_1(l)$. This will place the interaction point within V. Else, sample free path l from distribution $f_2(l)$.
3. Multiply particle weight by w, given by Eq. (4.46).

Distributions $f_1(l)$ and $f_2(l)$ can be sampled by the inversion method. The sampling formula for $f_1(l)$ is

$$l = -\frac{1}{\sigma} \ln[1 - \gamma + \gamma \exp(-\sigma s)], \tag{4.47}$$

and for distribution $f_2(l)$ the formula is

$$l = -\frac{1}{\sigma} \ln \gamma + s. \tag{4.48}$$

In an alternative, and more common version of the forced collisions technique, at the surface of the volume, the particle is split in two: one particle interacts within the volume (distribution $f_1(l)$), and the other particle starts at a point $l > s$ outside the volume (distribution $f_2(l)$). The bias is compensated by weights (see, for example, Abe et al. 2011). We have presented a more general version of the technique where the user can adjust the frequency of collisions in the volume of interest by

choosing an appropriate value of the collision probability p_c. In other words, p_c is an optimization parameter that can be adjusted to maximize the performance of a Monte Carlo code. With $p_c = 1/2$ our version of the technique is equivalent in statistical sense to this particle splitting algorithm. The difference is that instead of splitting each particle entering the volume, in our algorithm on average half of all particles (if $p_c = 1/2$) is forced to interact within the volume and the other half traverses the volume without an interaction.

4.4.5 *Exponential Transform*

This is yet another trajectory biasing technique. Some authors consider it standard for deep penetration problems, such as shielding calculations. In this technique the real total cross section, σ, is replaced by a modified cross section $\tilde{\sigma}$, which is a function of the polar angle θ, see Fig. 4.7.

This function is such that the modified cross section is small for particles traveling at small θ, i.e., for particles aimed at the detector. For particles traveling away from the detector, $\theta > \pi/2$, the cross section is large. A popular choice for the modified cross section is:

$$\tilde{\sigma} = \sigma - a \cdot \cos\theta = \sigma - a \cdot \mu. \tag{4.49}$$

The parameter a is called the stretching parameter. It is subject to constraint $0 < a < \sigma_{\min}$ that ensures the positivity of $\tilde{\sigma}$. The parameter can be a function of energy, but here we will present the formalism for a constant a. Adaptation to an energy dependent a is straightforward. The formalism is an example of the general trajectory biasing method introduced in Sect. 3.12.2. To implement this method, all we need is to find the weights Q_n that compensate for the trajectory bias. To do so, we will use Eqs. (3.208)–(3.209), but we now rewrite them with the modified cross section given by Eq. (4.49):

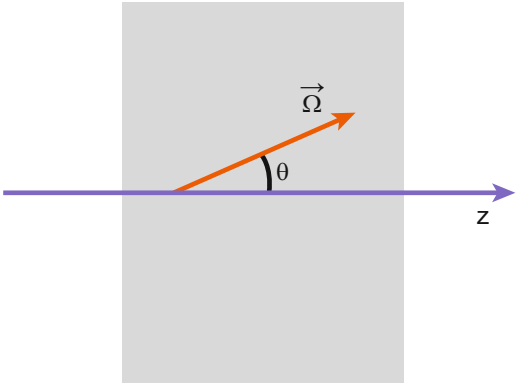

Fig. 4.7 Exponential transform

$$Q_{i+1} = Q_i \frac{\sigma\left(E_{i+1}\right)}{\tilde{\sigma}\left(E_{i+1}\right)} \exp\left\{\left[\tilde{\sigma}\left(E_{i+1}\right) - \sigma\left(E_{i+1}\right)\right] l_{i+1}\right\} \frac{\tilde{\sigma}\left(E_i\right)}{\sigma\left(E_i\right)}$$
$$= Q_i \frac{\sigma\left(E_{i+1}\right)}{\tilde{\sigma}\left(E_{i+1}\right)} \exp\left(-a\mu_{i+1} l_{i+1}\right) \frac{\tilde{\sigma}\left(E_i\right)}{\sigma\left(E_i\right)}. \tag{4.50}$$

$$Q_0 = \frac{\sigma\left(E_0\right)}{\tilde{\sigma}\left(E_0\right)} \frac{\Phi_0\left(x_0\right)}{\tilde{\Phi}_0\left(x_0\right)} = \frac{\sigma\left(E_0\right)}{\tilde{\sigma}\left(E_0\right)} \exp\left(-a\mu_0 l_0\right). \tag{4.51}$$

Because we did not bias the scattering cross sections, the ratio of differential scattering cross sections $\sigma_s(\cdots)/\tilde{\sigma}_s(\cdots)$ that is present in Eq. (3.208) is now equal to 1.

Notice that $\mu_{i+1} l_{i+1} = z_{i+1} - z_i$, where z_{i+1} and z_i are z-coordinates of the particle at the respective points of its trajectory. This leads to a significant simplification and saving of computing time. If a particle starts at $z = 0$, then its weight at the collision point k is simply

$$Q_k\left(x_0, x_1, \ldots, x_k\right) = \frac{\sigma\left(E_k\right)}{\tilde{\sigma}\left(E_k\right)} \exp\left(-az_k\right). \tag{4.52}$$

Hence, there is no need to calculate the weight for each particle at every step. Instead, the weight is calculated only for those particles that reach the detector and only when they do so. Interestingly enough, the weight does not explicitly depend on particle prior history, it is determined only by particle parameters at the current collision point.

4.4.6 Using Symmetry

The material included in this section was originally published in Kolchuzhkin and Vassiliev (1986). ©1986 Plenum Publishing Corporation. With permission of Springer.

In this technique particle trajectories are not biased. Instead, the source $S(x)$ is transformed into a point monodirectional source, and the detector $D(x)$ is modified so that the detector reading J remains exactly the same as it was in the original formulation of the problem. Algorithm efficiency improves because the modified detector is larger than the original and, therefore, more particles reach it. Additionally, sampling of the source distribution is simplified. Here, the formalism is presented for the maximum symmetry case. It is assumed that the medium is infinite and homogeneous. To apply the formalism to other symmetry types, the transformation operators used below, translation and rotation, need to be replaced by operators corresponding to the specific symmetry type. The formalism was developed by Kolchuzhkin and Vassiliev (1986).

Because the Boltzmann equation is linear, its solution can be expressed in terms of the Green's function $G(\vec{r}, \vec{\Omega}, E \mid \vec{r}\,', \vec{\Omega}\,', E')$:

$$\Phi\left(\vec{r},\vec{\Omega},E\right)=\int \mathrm{d}\vec{r}\,'\int \mathrm{d}\vec{\Omega}\,'\int \mathrm{d}E'G(\vec{r},\vec{\Omega},E\mid\vec{r}\,',\vec{\Omega}\,',E')S\left(\vec{r}\,',\vec{\Omega}\,',E'\right). \tag{4.53}$$

The Green's function $G(\vec{r},\vec{\Omega},E\mid\vec{r}\,',\vec{\Omega}\,',E')$ is the fluence at point $(\vec{r},\vec{\Omega},E)$ from a particle that originated at point $(\vec{r}\,',\vec{\Omega}\,',E')$. Accordingly, the equation for an observable takes the form

$$\begin{aligned}J=&\int \mathrm{d}\vec{r}\int \mathrm{d}\vec{\Omega}\int \mathrm{d}E\int \mathrm{d}\vec{r}\,'\int \mathrm{d}\vec{\Omega}\,'\int \mathrm{d}E\,'D\left(\vec{r},\vec{\Omega},E\right)\\&\times G(\vec{r},\vec{\Omega},E\mid\vec{r}\,',\vec{\Omega}\,',E')S\left(\vec{r}\,',\vec{\Omega}\,',E'\right).\end{aligned} \tag{4.54}$$

In a homogeneous infinite medium, the Green's function is invariant with respect to translations and rotations. Therefore,

$$G(\vec{r},\vec{\Omega},E\mid\vec{r}\,',\vec{\Omega}\,',E')=G(\vec{r}-\vec{r}\,',\vec{\Omega},E\mid\vec{0},\vec{\Omega}\,',E'), \tag{4.55}$$

and

$$G(\vec{r}-\vec{r}\,',\vec{\Omega},E|\vec{0},\vec{\Omega}\,',E')=G(\hat{R}_{[\vec{\Omega}',\vec{\Omega}_0]}\left(\vec{r}-\vec{r}\,'\right),\hat{R}_{[\vec{\Omega}',\vec{\Omega}_0]}\vec{\Omega},E|\vec{0},\vec{\Omega}_0,E'), \tag{4.56}$$

where $\hat{R}_{[\vec{\Omega}',\vec{\Omega}_0]}$ is the rotation operator in which vector $\vec{\Omega}'$ is superimposed on vector $\vec{\Omega}_0$ directed along the z-axis. If (θ',ϕ') are the polar and azimuthal angles of vector $\vec{\Omega}'$, then the corresponding rotation matrix is

$$\begin{aligned}\hat{R}_{[\vec{\Omega}',\vec{\Omega}_0]}&=\begin{pmatrix}\cos\theta' & 0 & -\sin\theta'\\ 0 & 1 & 0\\ \sin\theta' & 0 & \cos\theta'\end{pmatrix}\begin{pmatrix}\cos\phi' & \sin\phi' & 0\\ -\sin\phi' & \cos\phi' & 0\\ 0 & 0 & 1\end{pmatrix}\\&=\begin{pmatrix}\cos\theta'\cos\phi' & \cos\theta'\sin\phi' & -\sin\theta'\\ -\sin\phi' & \cos\phi' & 0\\ \sin\theta'\cos\phi' & \sin\theta'\sin\phi' & \cos\theta'\end{pmatrix}.\end{aligned} \tag{4.57}$$

The two matrices in the upper half of the equation represent two rotations: first, $\vec{\Omega}'$ was rotated about the z-axis by angle $-\phi'$, and then it was rotated about the y-axis by angle $-\theta'$. The matrix in the lower part of the equation is the product of the two matrices.

Next, we insert Eqs. (4.55) and (4.56) into Eq. (4.54), and replace the integration variables $\vec{r}$ and $\vec{\Omega}$ with new variables:

$$\vec{r}\,''=\hat{R}_{[\vec{\Omega}',\vec{\Omega}_0]}\left(\vec{r}-\vec{r}\,'\right);\qquad \vec{\Omega}\,''=\hat{R}_{[\vec{\Omega}',\vec{\Omega}_0]}\vec{\Omega}. \tag{4.58}$$

The Jacobian of this transformation of variables (translation and rotation) is equal to 1. The inverse transformation is

$$\vec{r} = \vec{r}\,' + \hat{R}^{-1}_{[\vec{\Omega}',\vec{\Omega}_0]}\vec{r}\,''; \qquad \vec{\Omega} = \hat{R}^{-1}_{[\vec{\Omega}',\vec{\Omega}_0]}\vec{\Omega}\,'', \tag{4.59}$$

where $\hat{R}^{-1}$ denotes the inverse operator. Then, the integral becomes

$$J = \int \mathrm{d}\vec{r}\,'' \int \mathrm{d}\vec{\Omega}\,'' \int \mathrm{d}E \int \mathrm{d}E' \tilde{D}\left(\vec{r}\,'', \vec{\Omega}\,'', E, E'\right) G\left(\vec{r}\,'', \vec{\Omega}\,'', E \mid \vec{0}, \vec{\Omega}_0, E'\right), \tag{4.60}$$

where

$$\tilde{D}\left(\vec{r}\,'', \vec{\Omega}\,'', E, E'\right) = \int \mathrm{d}\vec{r}\,' \int \mathrm{d}\vec{\Omega}\,' D\left(\vec{r}\,' + \hat{R}^{-1}_{[\vec{\Omega}',\vec{\Omega}_0]}\vec{r}\,'', \hat{R}^{-1}_{[\vec{\Omega}',\vec{\Omega}_0]}\vec{\Omega}\,'', E\right) \times S\left(\vec{r}\,', \vec{\Omega}\,', E'\right) \tag{4.61}$$

is the modified detector response function.

If in the source S, and detector D, angular and spatial dependence can be separated from the energy dependence, i.e.

$$S\left(\vec{r}, \vec{\Omega}, E\right) = S_E(E)\, S\left(\vec{r}, \vec{\Omega}\right); \tag{4.62}$$

$$D\left(\vec{r}, \vec{\Omega}, E\right) = D_E(E)\, D\left(\vec{r}, \vec{\Omega}\right), \tag{4.63}$$

then

$$\tilde{D}\left(\vec{r}, \vec{\Omega}, E, E'\right) = D_E(E)\, S_E\left(E'\right) \tilde{D}\left(\vec{r}, \vec{\Omega}\right), \tag{4.64}$$

where

$$\tilde{D}\left(\vec{r}, \vec{\Omega}\right) = \int \mathrm{d}\vec{r}\,' \int \mathrm{d}\vec{\Omega}\,' D\left(\vec{r}\,' + \hat{R}^{-1}_{[\vec{\Omega}',\vec{\Omega}_0]}\vec{r}, \hat{R}^{-1}_{[\vec{\Omega}',\vec{\Omega}_0]}\vec{\Omega}\right) S\left(\vec{r}\,', \vec{\Omega}\,'\right). \tag{4.65}$$

Inserting Eqs. (4.64) and (4.65) into Eq. (4.60) produces the following representation for the detector reading J:

$$J = \int \mathrm{d}\vec{r} \int \mathrm{d}\vec{\Omega} \int \mathrm{d}E \int \mathrm{d}E' D_E(E)\, \tilde{D}\left(\vec{r}, \vec{\Omega}\right) G\left(\vec{r}, \vec{\Omega}, E \mid \vec{0}, \vec{\Omega}_0, E'\right) S_E\left(E'\right). \tag{4.66}$$

This representation is the basis of our formalism for using symmetry of a problem to improve efficiency of Monte Carlo algorithms. Comparing this result with the original representation in Eq. (4.54), it can be seen that in a homogeneous infinite medium, the readings of a detector with sensitivity function $D_E(E)D(\vec{r}, \vec{\Omega})$, in the radiation field from a source $S_E(E)S(\vec{r}, \vec{\Omega})$ are equal to the readings of detector $D_E(E)\tilde{D}(\vec{r}, \vec{\Omega})$, in the radiation field of a point monodirectional source with energy spectrum $S_E(E)$. This formalism is applicable to other symmetry types. In that case,

the derivation of the modified detector response function should follow the same steps as we have just presented, starting with an appropriate form of Eqs. (4.55) and (4.56).

Example

The readings of a point isotropic detector located at point $\vec{r}^{\,*}$

$$D\left(\vec{r},\vec{\Omega}\right)=\delta\left(\vec{r}-\vec{r}^{\,*}\right) \tag{4.67}$$

in the field of a planar isotropic source located at $z=0$

$$S\left(\vec{r},\vec{\Omega}\right)=\frac{1}{4\pi}\delta\left(z\right) \tag{4.68}$$

is equal to the reading of a detector

$$\tilde{D}\left(\vec{r},\vec{\Omega}\right)=\begin{cases}\frac{1}{2r}, & \text{if } r>|z^*|\\ 0, & \text{otherwise}\end{cases} \tag{4.69}$$

in the field of the point monodirectional source

$$S_\delta\left(\vec{r},\vec{\Omega}\right)=\delta\left(\vec{r}\,\right)\delta\left(\vec{\Omega}-\vec{\Omega}_0\right). \tag{4.70}$$

Proof

In this example we ignore any energy dependence. All we need to do is to find the modified detector response function for the detector and the source given by Eqs. (4.67) and (4.68). To do so, we will use Eq. (4.65) which in this case has the following form:

$$\tilde{D}\left(\vec{r},\vec{\Omega}\right)=\int \mathrm{d}\vec{r}\,'\int \mathrm{d}\vec{\Omega}\,'\delta\left(\vec{r}\,'+\hat{R}^{-1}_{[\vec{\Omega}',\vec{\Omega}_0]}\vec{r}-\vec{r}^{\,*}\right)\frac{1}{4\pi}\delta\left(z'\right). \tag{4.71}$$

Operator $\hat{R}^{-1}_{[\vec{\Omega}',\vec{\Omega}_0]}$ only rotates the vector $\vec{r}$ and does not change its length. Hence, we can write $\hat{R}^{-1}_{[\vec{\Omega}',\vec{\Omega}_0]}\vec{r}=r\vec{\Omega}\,''$, and then replace the integration variable $\vec{\Omega}\,'$ with $\vec{\Omega}\,''$. We also change the order of integration to obtain:

$$\tilde{D}\left(\vec{r},\vec{\Omega}\right)=\frac{1}{4\pi}\int \mathrm{d}\vec{\Omega}\,''\int \mathrm{d}\vec{r}\,'\delta\left(\vec{r}\,'+r\vec{\Omega}\,''-\vec{r}^{\,*}\right)\delta\left(z'\right). \tag{4.72}$$

We can now integrate over x', y', and z'. The integration limits for each of the three variables are from $-\infty$ to $+\infty$. The result is:

$$\tilde{D}\left(\vec{r},\vec{\Omega}\right)=\frac{1}{4\pi}\int \mathrm{d}\vec{\Omega}''\delta\left(r\Omega_z''-z^*\right)=\frac{1}{4\pi}\int_0^{2\pi}\mathrm{d}\phi''\int_{-1}^{1}\mathrm{d}\mu''\delta\left(r\mu''-z^*\right). \tag{4.73}$$

Finally, in the second integral we replace the integration variable μ'' with $t=r\mu''$, and arrive at the expected result:

$$\tilde{D}\left(\vec{r},\vec{\Omega}\right)=\frac{1}{2r}\int_{-r}^{r}\mathrm{d}t\,\delta\left(t-z^*\right)=\begin{cases}\frac{1}{2r}, & \text{if } r>|z^*|,\\ 0, & \text{otherwise.}\end{cases} \tag{4.74}$$

□

This example shows how simple this technique can be. In the above proof, for example, we did not have to write the rotation matrix, or its inverse. Solving with Monte Carlo this problem in its original formulation, with an infinite planar source and a point detector, would be very challenging. In an algorithm copying the real process, particle trajectories would start at random positions on an infinite plane, and only a very small fraction of them reaching a small volume around the detector location $\vec{r}^*$ would contribute to the tally. Using symmetry of the problem results in a much more efficient algorithm. In the modified algorithm, all particles start at the origin and contribute to the tally each time they undergo a collision at a point outside a sphere of radius $|z^*|$ centered at the origin. Finally, it would be difficult, if possible, to derive the modified detector response function $\tilde{D}$, Eq. (4.74), solely from intuitive arguments, without using our formalism.

4.4.7 Correlated Sampling

Problem

For an observable J we need to calculate the difference $\Delta J = J_2 - J_1$ between two problems very similar in terms of the geometry, materials, sources, etc. If calculated with Monte Carlo, J_2 and J_1 are random numbers. If these two numbers are calculated independently, in separate calculations, then the variance of the difference ΔJ is approximately twice that of either J_1 or J_2, i.e.:

$$\mathrm{Var}\{\Delta J\}=\mathrm{Var}\{J_1\}+\mathrm{Var}\{J_2\}\approx 2\mathrm{Var}\{J_1\}. \tag{4.75}$$

Then, if the difference ΔJ is small, chances are, it cannot be estimated with a good accuracy.

Method

The idea of using correlated sampling comes from this equation

$$\mathrm{Var}\{\Delta J\} = \mathrm{Var}\{J_1\} + \mathrm{Var}\{J_2\} - 2\,\mathrm{Cov}\,(J_1, J_2)\,. \tag{4.76}$$

It is applicable when J_1 and J_2 are not statistically independent. It can be seen that $\mathrm{Var}\{\Delta J\}$ can be reduced if J_1 and J_2 are correlated and the covariance is positive. Given that the two problems are similar, the correlation is likely to be positive, in which case $\mathrm{Cov}(J_1, J_2) > 0$.

Algorithm

Correlation is achieved by calculating J_1 and J_2 on exactly the same trajectories. For example, trajectories are generated in a setup representing problem 1. Then all weights Q_n are equal to one when J_1 is calculated. The weights for J_2, however, need to be calculated because trajectories generated in the setup of problem 1 are biased, if used for solving problem 2. Trajectory biasing and calculations of the weights were discussed in Sect. 3.12.2. In that formalism the detector reading J is calculated as the expectation value of a random value ξ that is calculated as the following sum:

$$\xi = \sum_{n=0}^{\nu} Q_n h\,(x_n)\,. \tag{4.77}$$

To apply this formalism to calculate J_1 and J_2 on trajectories generated in the setup of problem 1, we will need to estimate the expectation values of these two random variables

$$\xi^{(1)} = \sum_{n=0}^{\nu} h\,(x_n)\,; \quad \xi^{(2)} = \sum_{n=0}^{\nu} Q_n h\,(x_n)\,. \tag{4.78}$$

It is important that the random states $x_0, x_1, \ldots x_\nu$ of a Markov chain in the two sums are exactly the same. Accordingly, to calculate the difference ΔJ, we will need to tally the following sum:

$$\xi_\Delta = \xi^{(2)} - \xi^{(1)} = \sum_{n=0}^{\nu} (Q_n - 1) h\,(x_n)\,. \tag{4.79}$$

4.5 Approximate Acceleration Techniques

Algorithm acceleration techniques are very common in Monte Carlo software. In fact, software not using at least one of the techniques presented in this section would be highly unusual. These techniques cannot be described as variance reduction, because they introduce systematic errors. These errors usually can be made as small as needed for a particular problem.

Condensed History Algorithms

Condensed history algorithms are very effective for solving problems involving charged particles. We discuss these algorithms in depth in the next chapter.

Range Rejection

This technique is usually used for electrons. In a voxelized geometry, when an electron is produced, its maximum range is calculated. If the range is smaller than the distance to the nearest voxel boundary, then its history is terminated. This introduces a systematic error, because this electron could produce a bremsstrahlung photon but its history was terminated before it could do so. Moreover, electron range is random and the actual range may exceed the calculated maximum range with some probability. This technique can also be used when geometry is not voxelized. In this case, electron trajectory is terminated, if the electron or its progenies cannot reach the detector, or contribute negligibly to detector reading.

Transport Cut-Offs and Production Thresholds

Transport cut-offs and production thresholds were discussed in Sect. 4.1. They are unavoidable in algorithms for transport of charged particles because in this case trajectories are essentially infinite. For example, an electron undergoes inelastic collisions and loses its kinetic energy, but it does not stop completely, nor is it absorbed. It is eventually thermalized and continues to move, now with the kinetic energy corresponding to the thermal energy of surrounding medium. The performance of algorithms for dose calculations from photon and electron sources is very sensitive to electron transport cut-offs and production thresholds. Cross sections for the production of secondary electrons are such that a large number

of low energy delta electrons is produced. If these cut-off and threshold energies are set too low, an avalanche of low energy electrons is produced, slowing down calculations significantly.

Kerma Approximation

In kerma approximation it is assumed that the energy transferred by neutral particles to matter is deposited locally, i.e., at the point of interaction, and that the absorbed dose is equal to kerma. This approximation is not generally applicable to photon beams in the megaelectron-volt range. However, therapeutic photon beams are polyenergetic and have a significant low-energy component. Then, for photons below a certain threshold energy, kerma approximation can be used, and secondary electrons produced by low energy photons do not need to be simulated.

We have named only a few acceleration techniques, chosen somewhat arbitrarily. Improving performance of Monte Carlo software has been a focus of extensive research in radiotherapy physics driven by the demand for fast and accurate software that can be used routinely for treatment planning. Abundant literature is available for more insight into approximate acceleration techniques.

References

Abe, S., Watanabe, Y., Niita, K., Sakamoto, Y.: Implementation of a forced collision method in the estimation of deposit energy distribution with the PHITS code. Prog. Nucl. Sci. Technol. **2**, 477–480 (2011)

Kolchuzhkin, A.M., Vassiliev, O.N.: Using symmetry properties to modify estimates in calculations of radiation fields by the Monte Carlo method. Atom. Energy **60**(4), 347–350 (1986)

Lewis, E.E., Miller, W.F.: Computational Methods of Neutron Transport. American Nuclear Society, La Grange Park (1993)

Vassiliev, O.N., Titt, U., Pönisch, F., Kry, S.F., Mohan, R., Gillin, M.T.: Dosimetric properties of photon beams from a flattening filter free clinical accelerator. Phys. Med. Biol. **51**(7), 1907–1917 (2006)

Wang, H., Vassiliev, O.N.: Microdosimetric characterisation of radiation fields for modelling tissue response in radiotherapy. Int. J. Cancer Ther. Oncol. **2**(1), 020116 (2014)

Chapter 5
Transport of Charged Particles

5.1 Overview

Charged particles that play an important role in radiotherapy include electrons, positrons, protons, and carbon ions. Other ions with relatively small atomic numbers are presently being considered as possible alternatives to proton and carbon-ion beams. Charged heavy recoils can also be produced in nuclear reactions. This effect can be significant and should be accounted for in dose calculations for hadrontherapy.

The most accurate Monte Carlo technique for charged particles involves event-by-event algorithms (see Chap. 6), where all interactions are simulated explicitly, and particles travel along a straight line with a constant energy between successive interactions. In other words, the same general algorithm is used for charged particles and neutral particles.

The drawback of this approach is the slowness of the algorithms due to the interaction cross sections being typically orders of magnitude larger for charged particles than for neutral particles. This is illustrated in Fig. 5.1 which shows tracks for three 10-MeV protons in water (Wang and Vassiliev 2014). Each dot in the figure represents an inelastic collision, either an ionization or an excitation. The tracks were generated with the Geant4-DNA software (Incerti et al. 2010; Bernal et al. 2015). The figure shows that the free path is of the order of only a few nanometers.

The standard solution to this problem is to use condensed history algorithms. In a first instance, let us consider only electrons. All interactions are classified as "catastrophic" or "soft," depending on the energy ΔE lost by an electron in a collision. If ΔE exceeds a threshold energy E_c, the collision is classified as catastrophic, and otherwise, as soft. The choice of threshold energy E_c depends on the desired level of accuracy. Choosing $E_c = 0$ maximizes the accuracy, but forces the algorithm to proceed "event-by-event." All catastrophic collisions are simulated explicitly. Multiple soft collisions are accounted for approximately by using a multiple scattering theory. Multiple scattering theories predict probability

O.N. Vassiliev, *Monte Carlo Methods for Radiation Transport*, Biological and Medical Physics, Biomedical Engineering, DOI 10.1007/978-3-319-44141-2_5

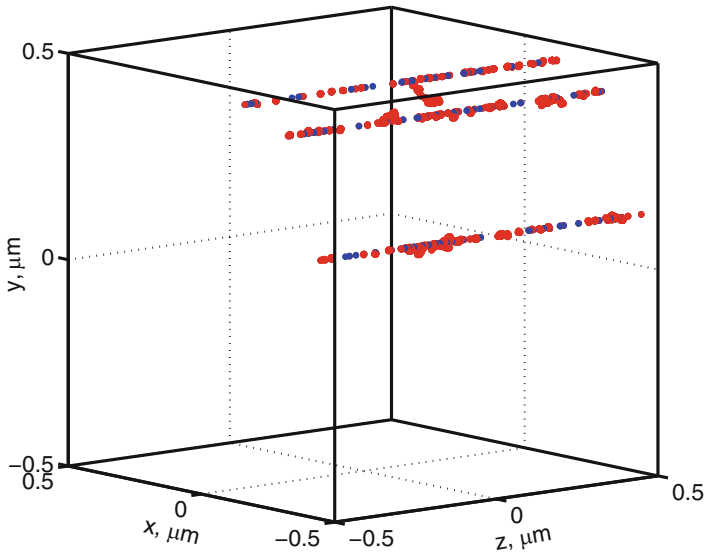

Fig. 5.1 Tracks of 10 MeV protons at a dose of approximately 2 Gy (Wang and Vassiliev 2014). Each dot represents an inelastic collision, ionization or excitation. Proton tracks are shown in *blue*, and delta electron tracks are in *red*. Used with permission

distributions of particle parameters after the particle has traveled a given distance. Such theories are applicable only to small distances. For this reason the electron trajectory between catastrophic collisions is generated step-by-step. The step size in a condensed history algorithm is an important parameter, chosen to optimize the balance between computing speed and accuracy. Examples of upper constraints on the step size are the energy lost by a particle and the magnitude of its lateral or angular deflection. The optimal step size depends on the particle energy and material properties. Several multiple scattering theories are introduced in this chapter.

Overview of a Typical Algorithm

1. Sample the initial phase coordinate of an electron, from the source distribution if it is emitted by the source or from differential cross sections, if it is a delta electron produced in a collision.
2. Sample a free path (distance) to a catastrophic collision.
3. Compare the free path to the step size.
4. If the free path is less then the step size, then:

   (a) Move the particle to the point of collision. An important difference from neutral particles is that, as an electron travels to the point of a catastrophic

collision, its parameters change in soft collisions. To sample the parameters of an electron when it arrives at the point of collision, multiple scattering theory is applied.

(b) Simulate the catastrophic collision.

(c) If secondary electrons are produced with energies exceeding E_c, their trajectories will need to be simulated. Simulations should start with the lowest energy particle. Parameters for all the other particles are saved to the computer memory.

(d) If secondary electrons with energies below E_c are produced, they do not need to be tracked, but the energy transferred to them must be accounted for.

5. If the free path exceeds the step size, then advance the particle position by a distance equal to the step size. Sample the particle parameters at the end of the step using multiple scattering theory. Make the next step, keeping in mind that the distance to the next catastrophic collision has already been sampled.

5.2 Energy Loss Models

5.2.1 The Continuous Slowing Down Approximation

We introduced the continuous slowing down approximation (CSDA) in Sect. 3.9.4. In this approximation a charged particles loses its energy continuously, according to the following formula:

$$\mathrm{d}E = -\beta\,(E)\,\mathrm{d}t, \tag{5.1}$$

where $\beta\,(E)$ is the stopping power, and $\mathrm{d}t$ is the distance traveled. Equation (5.1), in fact, defines the stopping power and can be solved to find the energy lost by a charged particle as it travels a given finite distance. The details are given in Sect. 3.9.4. In condensed history algorithms, in this equation the restricted stopping power $\beta_<(E)$ [defined in Eq. (3.130)] must be used. It accounts for energy losses only in soft collisions. Catastrophic collisions are simulated separately as discrete events. The main limitation of the CSDA is the assumption that the energy lost is not random. Radiation transport, however, is an inherently random process and energy loss fluctuations are always present. These fluctuations are called "energy straggling." They are important, and most Monte Carlo codes implement an energy straggling model. In some electron transport algorithms, however, fluctuations of energy loss in soft collisions are neglected to simplify the code and improve computing speed. In such cases energy straggling is partially accounted for through the sampling of catastrophic collisions. The three models of energy straggling presented in this section are relatively easy to implement and will not slow down computation significantly. They offer either an analytical solution or a solution that can be tabulated in a form convenient for sampling.

5.2.2 The Gaussian Model of Energy Straggling

The Gaussian model is a straightforward extension of the CSDA. In Sect. 3.9.4 we derived the collision integral within the CSDA, using a Taylor expansion of the product $\sigma_s(E', \Delta E)\Phi(E')$, Eq. (3.115). The expansion retained only the first two terms. The Gaussian model is derived by simply including the third term, quadratic in ΔE, in this expansion:

$$\sigma_s(E', \Delta E)\Phi(E') \approx \sigma_s(E, \Delta E)\Phi(E) + \frac{\partial}{\partial E}\left[\sigma_s(E, \Delta E)\Phi(E)\right]\Delta E + \frac{\partial^2}{\partial E^2}\left[\sigma_s(E, \Delta E)\Phi(E)\right]\frac{(\Delta E)^2}{2}. \tag{5.2}$$

This obviously improves the accuracy, as compared to the CSDA. We now use this result to derive a form of the Boltzmann equation. We then solve it to find the energy distribution of charged particles after they have traveled a given distance, t. In the context of condensed history algorithms, t is the step size. In doing so, we follow the methods developed by Landau (1944), and Vavilov (1957). First, we substitute Eq. (5.2) in the collision integral and integrate over ΔE. This step is very similar to Eq. (3.115):

$$\begin{aligned}\int_E^\infty \mathrm{d}E'\sigma_s\left(E' \to E\right)\Phi\left(E'\right) \approx & \int_0^\infty \sigma_s(E, \Delta E)\Phi\left(E\right)\mathrm{d}\left(\Delta E\right) \\ & + \frac{\partial}{\partial E}\int_0^\infty \left[\sigma_s(E, \Delta E)\Phi\left(E\right)\right]\left(\Delta E\right)\mathrm{d}\left(\Delta E\right) \\ & + \frac{\partial^2}{\partial E^2}\int_0^\infty \left[\sigma_s(E, \Delta E)\Phi\left(E\right)\right]\frac{(\Delta E)^2}{2}\mathrm{d}(\Delta E) \\ = & \ \sigma_s(E)\Phi\left(E\right) + \frac{\partial}{\partial E}\left[\beta(E)\Phi(E)\right] \\ & + \frac{1}{2}\frac{\partial^2}{\partial E^2}\left[\gamma\left(E\right)\Phi\left(E\right)\right],\end{aligned} \tag{5.3}$$

where $\gamma(E)$ is the mean square stopping power:

$$\gamma\left(E\right) = \int_0^\infty \sigma_s(E, \Delta E)\left(\Delta E\right)^2\mathrm{d}\left(\Delta E\right). \tag{5.4}$$

We assume that interaction parameters σ_s, β, and γ are all energy independent. A practical implication of this assumption is that it introduces an upper constraint on the step size. This constraint is usually expressed in terms of the energy lost. We now use this simplified form of the collision integral in the Boltzmann equation expressed in the Lagrangian form, Eq. (3.61), because the equation is written in

terms of the distance t traveled by the particle. We leave out the angular variables in order to focus on energy losses, and, for brevity, the initial phase coordinate of the particle $(\vec{r}_0, \vec{\Omega}_0, E_0)$. We also integrate the Boltzmann equation with respect to the spatial variables (x, y, z). This integration eliminates the streaming operator $(\vec{\Omega} \cdot \vec{\nabla})$ because by applying the Gauss–Ostrogradsky theorem (Appendix B) this integral can be transformed into a surface integral. We place the surface at infinity, where the fluence is zero. Finally, to simplify the algebra, we use $\sigma = \sigma_s$. The resulting form of the Boltzmann equation is as follows:

$$\frac{\partial}{\partial t}\Phi(t,E) = \beta\frac{\partial}{\partial E}\Phi(t,E) + \frac{\gamma}{2}\frac{\partial^2}{\partial E^2}\Phi(t,E). \tag{5.5}$$

A relatively simple method for solving Eq. (5.5) is based on the bilateral Laplace transform in energy. In this case, it is more convenient to use the variable $Q = E_0 - E$ instead of E. Note the difference between variables ΔE and Q. ΔE is the energy lost in one collision, whereas Q is the total energy lost by the particle as it travels a given distance. Replacing E with Q does not change Eq. (5.5) except for the sign of the first derivative:

$$\frac{\partial}{\partial E}\Phi(t,E) = -\frac{\partial}{\partial Q}\Phi(t,Q) \tag{5.6}$$

The initial condition for this equation is that at $t = 0$ the energy loss Q is zero:

$$\Phi(0,Q) = \delta(Q). \tag{5.7}$$

Because Eq. (5.5) is a second-order differential equation with respect to Q, two boundary conditions for $\Phi(t, Q)$ are also needed. These conditions are:

$$\Phi(t,\pm\infty) = \left[\frac{\partial}{\partial Q}\Phi(t,Q)\right]_{Q=\pm\infty} = 0. \tag{5.8}$$

The Laplace transformation of Eq. (5.5) with respect to Q is straightforward [see Abramowitz and Stegan (1964)]. We therefore proceed directly to the equation for the Laplace transform of the fluence, $\Phi(t, p)$:

$$\frac{\partial}{\partial t}\Phi(t,p) = -\beta p\Phi(t,p) + \frac{\gamma}{2}p^2\Phi(t,p). \tag{5.9}$$

The solution for the Laplace transform is then

$$\Phi(t,p) = \exp\left[\left(-\beta p + \frac{\gamma}{2}p^2\right)t\right], \tag{5.10}$$

where we applied the boundary condition for the Laplace transform, $\Phi(0, p) = 1$, that follows from Eq. (5.7).

The remaining task is to perform the inverse Laplace transform by calculating the following integral:

$$\Phi(t,Q) = \frac{1}{2\pi i}\int_{-i\infty}^{+i\infty} \mathrm{e}^{pQ}\Phi(t,p)\,\mathrm{d}p = \frac{1}{2\pi i}\int_{-i\infty}^{+i\infty} \exp\left[pQ + \left(-\beta p + \frac{\gamma}{2}p^2\right)t\right]\mathrm{d}p. \tag{5.11}$$

We first rewrite the exponent as follows:

$$pQ - \beta pt + \frac{\gamma}{2}p^2 t = \left(\frac{Q-\beta t}{\sqrt{2\gamma t}} + p\sqrt{\frac{\gamma t}{2}}\right)^2 - \frac{(Q-\beta t)^2}{2\gamma t}. \tag{5.12}$$

The integral then becomes:

$$\Phi(t,Q) = \exp\left[-\frac{(Q-\beta t)^2}{2\gamma t}\right]\frac{1}{2\pi i}\int_{-i\infty}^{+i\infty}\exp\left(\frac{Q-\beta t}{\sqrt{2\gamma t}} + p\sqrt{\frac{\gamma t}{2}}\right)^2\mathrm{d}p. \tag{5.13}$$

This is further simplified by the substitution:

$$w = \frac{1}{i}\left(\frac{Q-\beta t}{\sqrt{2\gamma t}} + p\sqrt{\frac{\gamma t}{2}}\right), \tag{5.14}$$

leading to

$$\Phi(t,Q) = \exp\left[-\frac{(Q-\beta t)^2}{2\gamma t}\right]\frac{1}{2\pi}\sqrt{\frac{2}{\gamma t}}\int_{-\infty}^{+\infty}\mathrm{e}^{-w^2}\mathrm{d}w. \tag{5.15}$$

The remaining integral over w equals $\sqrt{\pi}$, and the final result is the normal distribution:

$$\Phi(t,Q) = \frac{1}{\sqrt{2\pi\gamma t}}\exp\left[-\frac{(Q-\beta t)^2}{2\gamma t}\right]. \tag{5.16}$$

According to this result, the average energy lost is βt, which is consistent with the continuous slowing down approximation, assuming that the stopping power β was constant while the particle traveled over the distance t. The width of the distribution increases with increasing t. The standard deviation is $\sqrt{\gamma t}$. The limit $\gamma \to 0$ produces a delta function: $\Phi(t,Q) = \delta(Q-\beta t)$. Similarly, the limit $t \to 0$ yields $\Phi(t,Q) = \delta(Q)$, consistent with the initial condition Eq. (5.7), as expected.

This result is applicable only to short travel distances, where $Q \ll E_0$. Also, because Q must be positive, the probability of a negative Q should be small. Equation (5.16) suggests a connection to the central limit theorem. Indeed, the total

energy loss Q is the sum of energies lost in individual inelastic collisions. This observation implies a lower bound on the step size z, it should be sufficiently large that the average number of inelastic collisions is much greater than one.

5.2.3 The Landau Model

In this model introduced by Landau (1944), the differential scattering cross section $\sigma_s(E' \to E)$ is approximated by the Rutherford formula (Rutherford 1911; Nikjoo et al. 2008). A limitation of the formula is that it describes the scattering of a charged particle on a free target particle. Here, however, we are primarily concerned with energy losses in inelastic interactions with atomic electrons. Hence, the target particle is not free but bound to an atom. For this reason, in this model, the Rutherford formula is applied only to collisions where the energy loss ΔE is much greater than the binding energy of atomic electrons.

We start with the Boltzmann equation in a form similar to Eq. (5.5), but with the collision integral written exactly:

$$\frac{\partial}{\partial t}\Phi\left(t,E\right)+\sigma\left(t,E\right)=\int_0^\infty \sigma_s\left(t,E+\Delta E,\Delta E\right)\Phi\left(t,E+\Delta E\right)\mathrm{d}(\Delta E). \tag{5.17}$$

We apply this equation to particles traveling distances sufficiently small to justify the assumption that the total energy Q the particle loses over these distances is much less than its initial energy E_0. We can then assume that $\sigma_s(E+\Delta E,\Delta E)\approx\sigma_s(E_0,\Delta E)$ and, for brevity, leave out E_0. The next step is the same as in the previous section. We assume that $\sigma(t,E)=\sigma_s(t,E)=\sigma_s(E)\approx\sigma_s(E_0)$, and that the fluence at $t=0$ is monoenergetic, and we switch from the variable E to $Q=E_0-E$. Then, Eq. (5.17) becomes:

$$\frac{\partial}{\partial t}\Phi\left(t,Q\right)+\sigma_s\Phi\left(t,Q\right)=\int_0^\infty \sigma_s\left(\Delta E\right)\Phi\left(t,Q-\Delta E\right)\mathrm{d}(\Delta E). \tag{5.18}$$

We write $Q-\Delta E$ as the fluence argument in the collision integral because it accounts for particles that, prior to a collision, had lost an energy $Q-\Delta E$, and then, in the collision lose a further ΔE, thus contributing to the fluence $\Phi(Q)$ of particles losing a total energy Q. Here, following Landau (1944), we set the upper limit of integration to infinity, because $\Phi(t,Q)=0$ when $Q<0$.

We next apply the Laplace transform, with respect to Q, to both sides of Eq. (5.18), which yields an equation for the Laplace transform of the fluence, $\Phi(t,p)$:

$$\frac{\partial}{\partial t}\Phi\left(t,p\right)+\sigma_s\Phi\left(t,p\right)=\int_0^\infty \sigma_s\left(\Delta E\right)\Phi\left(t,p\right)\mathrm{e}^{-p\Delta E}\mathrm{d}(\Delta E);\ \ t>0, \tag{5.19}$$

where we used the property that a shift in the argument of a function (e.g., $\Phi(Q - \Delta E)$ in the present case) translates as an exponent in the Laplace transform (Arfken et al. 2013). From Eq. (5.19) we find:

$$\begin{aligned}\Phi(t,p) &= \exp\left[t\int_0^\infty \sigma_s(\Delta E)\,\mathrm{e}^{-p\Delta E}\mathrm{d}(\Delta E) - \sigma_s t\right] \\ &= \exp\left[-t\int_0^\infty \sigma_s(\Delta E)\left(1-\mathrm{e}^{-p\Delta E}\right)\mathrm{d}(\Delta E)\right].\end{aligned} \tag{5.20}$$

The solution $\Phi(t, Q)$ of Eq. (5.18) can, in principle, be found by calculating the inverse Laplace transform of Eq. (5.20). However, to simplify the right-hand side of Eq. (5.20), we further assume that there exists an energy ϵ_1 that is much greater than the average binding energy of the atomic electrons, while being sufficiently small to justify the approximation $\exp(-p\Delta E) \approx 1 - p\Delta E$, when $\Delta E < \epsilon_1$. We then split the integral in Eq. (5.20) into two parts: one with $\Delta E < \epsilon_1$ and the other with $\Delta E > \epsilon_1$. The first part is simple:

$$\int_0^{\epsilon_1} \sigma_s(\Delta E)\left(1-\mathrm{e}^{-p\Delta E}\right)\mathrm{d}(\Delta E) \approx \int_0^{\epsilon_1} \sigma_s(\Delta E)\,p\Delta E\mathrm{d}(\Delta E) = p\beta_<, \tag{5.21}$$

where $\beta_<$ is the restricted stopping power, i.e., the stopping power due to collisions with energy transfers less than ϵ_1. In the second part ($\Delta E > \epsilon_1$) we use the Rutherford formula for the scattering cross section (Rutherford 1911; Nikjoo et al. 2008):

$$\sigma_s(\Delta E) = \frac{A}{(\Delta E)^2}. \tag{5.22}$$

The coefficient A depends on the charge and energy of the incident particle, as well as on the material properties. The integral for $\Delta E > \epsilon_1$ is then integrated by parts twice, which reduces it to an integral representation of the Euler constant γ:

$$\gamma = -\int_0^\infty \mathrm{e}^{-x}\ln x\mathrm{d}x \approx 0.5772. \tag{5.23}$$

The integration is performed as follows:

$$\begin{aligned}\int_{\epsilon_1}^\infty \sigma_s(\Delta E)\left(1-\mathrm{e}^{-p\Delta E}\right)\mathrm{d}(\Delta E) &= A\int_{\epsilon_1}^\infty \frac{1-\mathrm{e}^{-p\Delta E}}{(\Delta E)^2}\mathrm{d}(\Delta E) \\ &= \frac{A}{\epsilon_1}\left(1-\mathrm{e}^{-p\epsilon_1}\right) + A\int_{\epsilon_1}^\infty \frac{p}{\Delta E}\mathrm{e}^{-p\Delta E}\mathrm{d}(\Delta E)\end{aligned}$$

$$= \frac{A}{\epsilon_1}\left(1 - \mathrm{e}^{-p\epsilon_1}\right) - Ap \ln\left(\epsilon_1 p\right)\mathrm{e}^{-p\epsilon_1} + Ap \int_{\epsilon_1 p}^{\infty} \ln x \, \mathrm{e}^{-x} \mathrm{d}x. \tag{5.24}$$

Finally, recalling that ϵ_1 is small, we write the last formula in asymptotic form, keeping only the terms that do not vanish in the limit $\epsilon_1 p \to 0$:

$$\int_{\epsilon_1}^{\infty} \sigma_s(\Delta E)\left(1 - \mathrm{e}^{-p\Delta E}\right) \mathrm{d}(\Delta E) \approx Ap\left[1 - \ln\left(\epsilon_1 p\right) - \gamma\right]. \tag{5.25}$$

This last step is somewhat heuristic and warrants an analysis of uncertainties. We defer such a discussion to the next section, where this result can be compared to a more general energy straggling model.

The fluence spectrum is now found by performing the inverse Laplace transform:

$$\Phi\left(t, Q\right) = \frac{1}{2\pi i} \int_{c-i\infty}^{c+i\infty} \Phi\left(t, p\right) \mathrm{e}^{pQ} \mathrm{d}p. \tag{5.26}$$

This integration cannot be done analytically. However, the solution can be expressed in terms of a universal function $\phi(\lambda)$ of a dimensionless argument λ. Using Eqs. (5.20), (5.21), (5.25), and substitution $u = Apt$, the solution can be represented in the following form

$$\Phi\left(t, Q\right) = \frac{\phi\left(\lambda\right)}{tA}, \tag{5.27}$$

where

$$\phi\left(\lambda\right) = \frac{1}{2\pi i} \int_{c-i\infty}^{c+i\infty} \exp\left(u \ln u + \lambda u\right) \mathrm{d}u, \tag{5.28}$$

and

$$\lambda = \frac{Q}{tA} - \frac{\beta_<}{A} - 1 - \ln\frac{tA}{\epsilon_1} + \gamma. \tag{5.29}$$

The integral in Eq. (5.28) can be transformed into a form that is more convenient for calculations. First, we integrate along the imaginary axis by setting $c = 0$ in the integral limits. Then, we split the integral into a sum of two parts: an integral from $-i\infty$ to 0 and from 0 to $+i\infty$. In the first integral we substitute $u = -iy$, and in the second, $u = iy$. Using the property of the logarithm

$$\ln\left(\pm iy\right) = \ln y \pm i\frac{\pi}{2}, \tag{5.30}$$

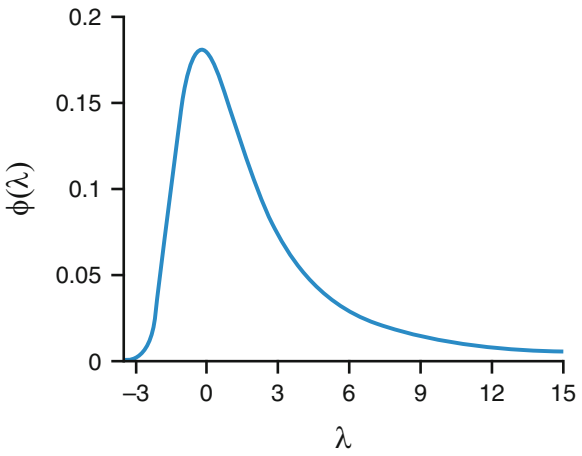

Fig. 5.2 The function $\phi(\lambda)$ of the Landau model of energy straggling, Eq. (5.31)

the two integrals can then be combined to give the following representation of the function ϕ:

$$\phi(\lambda) = \frac{1}{\pi}\int_0^\infty \mathrm{e}^{-(\pi/2)y}\cos(y\ln y + \lambda y)\mathrm{d}y. \tag{5.31}$$

A graph of $\phi(\lambda)$ is shown in Fig. 5.2.

5.2.4 *The Vavilov Model*

This model developed by Vavilov (1957) is more general than the two previous models. We will show that both the Gaussian and the Landau models are in fact special cases of the Vavilov model.

We start by inserting the Laplace transform of fluence $\Phi(t,p)$ given by Eq. (5.20) into the formula for inverse transform, Eq. (5.26):

$$\Phi(t,Q) = \frac{1}{2\pi i}\int_{c-i\infty}^{c+i\infty}\exp\left\{pQ - t\int_0^{(\Delta E)_{\max}}\sigma_s(\Delta E)\left(1-\mathrm{e}^{-p\Delta E}\right)\mathrm{d}(\Delta E)\right\}\mathrm{d}p. \tag{5.32}$$

The difference between this equation and the corresponding equation in the Landau model is that the infinite upper integration limit in the exponent is replaced by $(\Delta E)_{\max}$, the maximum energy that can be transferred in a single collision. Assuming that the incident particle is much heavier than the target electron, we have

$$(\Delta E)_{\max} = \frac{2m_e v^2}{1 - v^2/c^2}, \tag{5.33}$$

where v is the particle velocity and c is the speed of light. Another difference—indeed an improvement—is that the Rutherford formula is written in a relativistic form:

$$\sigma_s(\Delta E) = \frac{A}{(\Delta E)^2}\left[1 - \left(\frac{v}{c}\right)^2 \frac{\Delta E}{(\Delta E)_{\max}}\right], \tag{5.34}$$

As indicated in the preceding section, the Rutherford formula applies only when the energy transfer is much greater than the binding energy of atomic electrons. Vavilov (1957) proposed an elegant method for addressing this limitation, modifying the exponent in the integral in Eq. (5.32) as follows:

$$\{\ldots\} = p(Q - \beta t) - t\int_0^{(\Delta E)_{\max}} \sigma_s(\Delta E)\left(1 - \mathrm{e}^{-p\Delta E} - p\Delta E\right)\mathrm{d}(\Delta E), \tag{5.35}$$

where β is the total stopping power defined in Eq. (3.116). This simple modification eliminates the divergence of the integral over ΔE in Eq. (5.32) at $\Delta E \to 0$. It also reduces the contribution to the integral from collisions with a small ΔE relative to collisions with large ΔE, which justifies the use of the Rutherford formula. We can now insert Eq. (5.34) into Eq. (5.35) and calculate the integral. The integration is rather lengthy, but we will explain every step. We start by splitting the integral into two parts, I_1 and I_2:

$$\int_0^{(\Delta E)_{\max}} \frac{A}{(\Delta E)^2}\left[1 - \left(\frac{v}{c}\right)^2 \frac{\Delta E}{(\Delta E)_{\max}}\right]\left(1 - \mathrm{e}^{-p\Delta E} - p\Delta E\right)\mathrm{d}(\Delta E) = I_1 - I_2, \tag{5.36}$$

where

$$I_1 = \int_0^{(\Delta E)_{\max}} \frac{A}{(\Delta E)^2}\left(1 - \mathrm{e}^{-p\Delta E} - p\Delta E\right)\mathrm{d}(\Delta E), \tag{5.37}$$

and

$$I_2 = \int_0^{(\Delta E)_{\max}} \frac{A}{(\Delta E)^2}\left(\frac{v}{c}\right)^2 \frac{\Delta E}{(\Delta E)_{\max}}\left(1 - \mathrm{e}^{-p\Delta E} - p\Delta E\right)\mathrm{d}(\Delta E). \tag{5.38}$$

Then, we integrate I_1 by parts:

$$I_1 = \frac{-A}{(\Delta E)_{\max}}\left[1 - \mathrm{e}^{-p(\Delta E)_{\max}} - p(\Delta E)_{\max}\right] - Ap\int_0^{(\Delta E)_{\max}} \frac{1 - \mathrm{e}^{-p\Delta E}}{\Delta E}\mathrm{d}(\Delta E)\,. \tag{5.39}$$

The remaining integral can be expressed in terms of the Euler constant γ, Eq. (5.23), and the exponential integral Ei, defined as follows (Arfken et al. 2013):

$$\mathrm{Ei}(x) = -\int_{-x}^{\infty} \frac{\mathrm{e}^{-w}}{w}\mathrm{d}w. \tag{5.40}$$

To achieve this, we again, split the integral into two:

$$\int_0^{(\Delta E)_{\max}} \ldots \ \mathrm{d}(\Delta E) = \lim_{T\to+\infty} \left\{ \int_0^T \ldots \ \mathrm{d}(\Delta E) - \int_{(\Delta E)_{\max}}^T \ldots \ \mathrm{d}\,(\Delta E) \right\}. \tag{5.41}$$

The first integral in the brackets is integrated by parts, with the integration variable changed to $w = p\Delta E$:

$$\int_0^{pT} \frac{1-\mathrm{e}^{-w}}{w}\mathrm{d}w = \ln(pT)\left(1-\mathrm{e}^{-pT}\right) - \int_0^{pT} \ln w \ \mathrm{e}^{-w}\mathrm{d}w. \tag{5.42}$$

The second integral is simply:

$$\int_{p(\Delta E)_{\max}}^{pT} \frac{1-\mathrm{e}^{-w}}{w}\mathrm{d}w = \ln(pT) - \ln\left[p\,(\Delta E)_{\max}\right] - \int_{p(\Delta E)_{\max}}^{pT} \frac{\mathrm{e}^{-w}}{w}\mathrm{d}w. \tag{5.43}$$

Subtracting Eq. (5.43) from Eq. (5.42) and taking the limit $T \to +\infty$, we obtain

$$\int_0^{(\Delta E)_{\max}} \frac{1-\mathrm{e}^{-p\Delta E}}{\Delta E}\mathrm{d}\,(\Delta E) = \gamma + \ln\left[p\,(\Delta E)_{\max}\right] - \mathrm{Ei}\left[-p\,(\Delta E)_{\max}\right], \tag{5.44}$$

where γ is the Euler constant, see Eq. (5.23). The integral I_2 is simple. A part of it is integrated analytically and the rest reduces to the same integral as above:

$$I_2 = \frac{A}{(\Delta E)_{\max}}\left(\frac{v}{c}\right)^2 \left\{ \int_0^{(\Delta E)_{\max}} \frac{1-\mathrm{e}^{-p\Delta E}}{\Delta E}\mathrm{d}(\Delta E) - p\,(\Delta E)_{\max} \right\}. \tag{5.45}$$

We next enter I_1 and I_2 into Eq. (5.35), and then into the inverse transform formula, Eq. (5.32), where we change the integration variable from p to $q = p(\Delta E)_{\max}$. This yields a rather complicated expression for the Laplace transform of the fluence:

$$\Phi(t,Q) = \frac{\exp\left[k\left(1+\gamma v^2/c^2\right)\right]}{2\pi i\,(\Delta E)_{\max}} \times \int_{c-i\infty}^{c+i\infty} \exp\left\{ q\lambda_1 + k\left[\left(q+\left(\frac{v}{c}\right)^2\right)(\ln q - \mathrm{Ei}(-q)) - \mathrm{e}^{-q}\right]\right\}\mathrm{d}q, \tag{5.46}$$

where

$$\lambda_1 = \frac{Q-\beta t}{(\Delta E)_{\max}} - k\left[1+\left(\frac{v}{c}\right)^2 - \gamma\right], \tag{5.47}$$

and

$$k = \frac{At}{(\Delta E)_{\max}}. \tag{5.48}$$

To make sense of this result, let us first consider two extreme cases: $k \ll 1$ and $k \gg 1$. We will show that in the first case, the result coincides with the Landau distribution, and in the second case, with the normal distribution.

Recalling that the Landau model is nonrelativistic, we set v/c to zero and then replace the integration variable q with $u = kq$. This new integration variable is similar to the integration variable in the Landau model, Eq. (5.28). The exponent in Eq. (5.46) then becomes:

$$\left(\frac{u}{k}\right)\frac{Q-\beta t}{(\Delta E)_{\max}} - u\,(1-\gamma) + u\ln\frac{u}{k} - u\mathrm{Ei}\left(-\frac{u}{k}\right) - k\mathrm{e}^{-u/k}. \tag{5.49}$$

In the limit $k \to 0$ the last two terms cancel out, because of the asymptotic behavior of the exponent integral for large arguments (Abramowitz and Stegan 1964):

$$\mathrm{Ei}\,(x) \approx \frac{\mathrm{e}^x}{x}. \tag{5.50}$$

Comparing Eq. (5.49) with the exponent in the Landau model, Eq. (5.28), we can see that we will achieve the desired result, that is Landau's distribution, if we can show that

$$\left(\frac{u}{k}\right)\frac{Q-\beta t}{(\Delta E)_{\max}} - u\,(1-\gamma) - u\ln k = \lambda u. \tag{5.51}$$

To do so, we note that the total stopping power β on the left-hand side of Eq. (5.51) is the sum of the stopping powers from soft and hard collisions, $\beta = \beta_< + \beta_>$. The latter contribution is easily calculated:

$$\beta_> = \int_{\epsilon_1}^{(\Delta E)_{\max}} \frac{A}{(\Delta E)^2}\,(\Delta E)\,\mathrm{d}\,(\Delta E) = A\ln\left[\frac{(\Delta E)_{\max}}{\epsilon_1}\right]. \tag{5.52}$$

Using this result, and recalling the definitions of λ, Eq. (5.29), and k, Eq. (5.48), the proof of Eq. (5.51) is straightforward.

To summarize, we have shown that the Vavilov distribution coincides with the Landau distribution, if the particles are nonrelativistic, and k is small. The parameter k is proportional to the distance traveled, t, and hence the Landau distribution applies for small travel distances where the condition $k \ll 1$ is satisfied.

The Gaussian model can be recovered for a large step size, t, i.e., when $k \gg 1$. To show this, we will calculate the integral on the right-hand side of Eq. (5.35), using a different method than above. We will use the nonrelativistic form of the

Rutherford formula, Eq. (5.22), and a Taylor series expansion of the exponent ($e^y = \sum_{n=0}^{\infty} y^n/n!$):

$$\int_0^{(\Delta E)_{\max}} \frac{A}{(\Delta E)^2}\left(1 - e^{-p\Delta E} - p\Delta E\right) d(\Delta E) = |\text{substitution } y = -p\Delta E|$$

$$= -Ap\int_0^{-p(\Delta E)_{\max}} \frac{1}{y^2}\left[1 + y - \sum_{n=0}^{\infty}\frac{y^n}{n!}\right] dy$$

$$= Ap\sum_{n=2}^{\infty}\frac{1}{n!}\frac{[-p\,(\Delta E)_{\max}]^{n-1}}{n-1}. \tag{5.53}$$

Let us retain only the first two terms in the sum, $n = 2, 3$, and insert the above result in Eq. (5.35). The equation then takes the form

$$\{\ldots\} = p\,(Q - \beta t) + \frac{1}{2}A\,(\Delta E)_{\max}\,tp^2 - \frac{1}{12}A\,(\Delta E)^2_{\max}\,tp^3. \tag{5.54}$$

To find the energy loss spectrum $\Phi(t, Q)$, we need to perform the inverse Laplace transform, as given by Eq. (5.32). This can be done analytically (Vavilov 1957). We, however, will discuss the result given by Eq. (5.54) only qualitatively. If in Eq. (5.54) we retain only terms linear in p, then the spectrum will be a delta function, $\Phi(t, Q) \propto \delta(Q - \beta t)$. If we retain only terms linear and quadratic in p, then the spectrum will be Gaussian, as in Eq. (5.16). Including also the term cubic in p, will add asymmetry to the Gaussian curve. As the step size t, increases, the maximum of the energy distribution, $Q_{\max}$ shifts towards higher values, so that the ratio $Q_{\max}/(\beta t)$ remains approximately constant. Keeping this in mind, we change the integration variable in the inverse transform integral in Eq. (5.32) from p to $w = pQ$, and then write Eq. (5.54) as

$$\{\ldots\} = w - w\left(\frac{\beta t}{Q}\right) + w^2\frac{A\,(\Delta E)_{\max}}{2\beta}\left(\frac{\beta t}{Q}\right)^2\frac{1}{\beta t} - w^3\frac{A\,(\Delta E)^2_{\max}}{12\beta}\left(\frac{\beta t}{Q}\right)^3\frac{1}{\beta^2 t^2}. \tag{5.55}$$

For a sufficiently large step size, t, the last term in Eq. (5.55) becomes negligible, and the energy loss distribution $\Phi(t, Q)$, indeed, becomes Gaussian.

In conclusion, when $k \ll 1$, we can use the Landau model, and when $k \gg 1$, the Gaussian model. When neither model applies, the full version of the Vavilov model should then be used.

The energy loss distribution in this full model is given by Eq. (5.46), which can be transformed into a more convenient form for calculations. The integration must be performed along the imaginary axis. However, this can be avoided by splitting the integral into two parts as shown below, and substituting $q = iy_1$ in the first integral, and $q = -iy_2$ in the second. Both y_1 and y_2 are real:

$$\int_{-i\infty}^{+i\infty} \dots \mathrm{d}q = \int_{0}^{+i\infty} \dots \mathrm{d}q + \int_{-i\infty}^{0} \dots \mathrm{d}q = i\int_{0}^{\infty} \dots \mathrm{d}y_1 + i\int_{0}^{\infty} \dots \mathrm{d}y_2. \tag{5.56}$$

The corresponding transformation of the integrand is straightforward, if we use Eq. (5.30) for the logarithm, and also recall that

$$\mathrm{Ei}\,(\pm \mathrm{i}y) = \mathrm{Ci}\,(y) \pm i\left[\mathrm{Si}\,(y) - \frac{\pi}{2}\right], \tag{5.57}$$

where Ci(y) and Si(y) are the cosine and sine integrals, respectively. Using Eqs. (5.30), (5.56), and (5.57) we can rewrite Eq. (5.46) somewhat more simply

$$\Phi\,(t,Q) = \frac{\exp\left[k\left(1+\gamma v^2/c^2\right)\right]}{\pi\,(\Delta E)_{\max}} \int_0^{\infty} \mathrm{e}^{kf_1(y)} \cos\left[\lambda_1 y + kf_2\,(y)\right]\mathrm{d}y, \tag{5.58}$$

where

$$f_1\,(y) = \left(\frac{v}{c}\right)^2 \left[\ln y - \mathrm{Ci}\,(y)\right] - y\mathrm{Si}\,(y) - \cos y, \tag{5.59}$$

and

$$f_2\,(y) = y\left[\ln y - \mathrm{Ci}\,(y)\right] + \left(\frac{v}{c}\right)^2 \mathrm{Si}\,(y) + \sin y. \tag{5.60}$$

The integral in Eq. (5.58) can be calculated numerically. For nonrelativistic particles ($v/c = 0$) it can be tabulated as a function of k and λ_1. Figure 5.3 illustrates the overall behavior of the full Vavilov solution. It shows distributions of energy loss as a function of λ_1, calculated using Eq. (5.58) for a 100 MeV proton and $k = 0.2$, 0.4, 0.6, 0.8, and 1. In the figure the fluence is multiplied by At, making

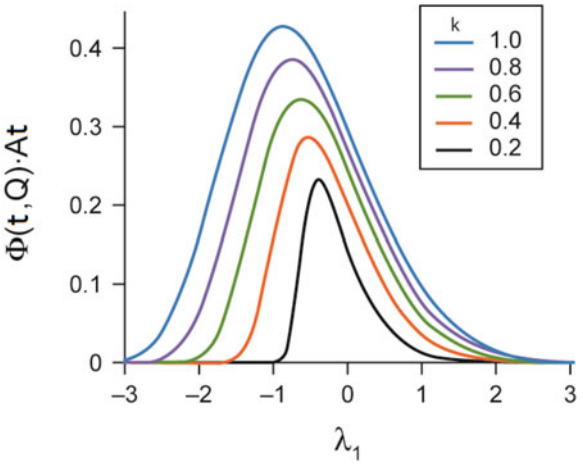

Fig. 5.3 Vavilov's distribution given by Eq. (5.58) for a 100 MeV proton and a range of k values as indicated

this result also applicable to other charged particles traveling with the same velocity as a 100 MeV proton. It can be seen that as k increases the distribution becomes more symmetric, resembling a normal distribution. At small k, as expected, its shape becomes similar to Landau's distribution shown in Fig. 5.2. However, it should be noted that the axes in Figs. 5.2 and 5.3 are different.

5.3 Models for the Angular Distribution

5.3.1 The Fokker–Planck Approximation

The Fokker–Planck form of the collision integral in the Boltzmann equation was derived in Sect. 3.9.6, Eqs. (3.141)–(3.143). In this section we will solve the equation, after making additional approximations. We will assume that the medium is homogeneous, and as the particle travels, its energy loss is negligible and angular deflection is small. The last two conditions can be satisfied by limiting the step size. As for the homogeneity, a standard method for modeling a heterogeneous material is to approximate it using a voxelized geometry. While the particle is located within a particular voxel, it travels in a homogeneous medium. If all these conditions are reasonably met, the Fokker–Planck approximation provides a very simple algorithm for sampling angular distributions.

We start with the Boltzmann equation in the Lagrangian form, Eq. (3.61). For brevity we will omit energy variables, and the initial phase coordinate of the particle $(\vec{r}_0, \vec{\Omega}_0, E_0)$. We will also assume that the only process is scattering, so that $\sigma = \sigma_s$. Then, the equation is

$$\frac{\partial}{\partial t}\Phi\left(\vec{r},\vec{\Omega},t\right) + \left(\vec{\Omega}\cdot\vec{\nabla}\right)\Phi\left(\vec{r},\vec{\Omega},t\right) + \sigma_s\Phi\left(\vec{r},\vec{\Omega},t\right) = \int \mathrm{d}\vec{\Omega}\,'\sigma_s\left(\vec{\Omega}\,'\cdot\vec{\Omega}\right)\Phi\left(\vec{r},\vec{\Omega}\,',t\right). \tag{5.61}$$

Next, we integrate Eq. (5.61) over $\vec{r}$. This integration eliminates the streaming operator, $(\vec{\Omega}\cdot\vec{\nabla})\Phi$, and replaces the fluence $\Phi(\vec{r},\vec{\Omega},t)$ with $\Phi(\vec{\Omega},t)$. The former fluence is the angular distribution of a particle that has traveled distance t and is located at point $\vec{r}$, the latter is the angular distribution of a particle that has traveled distance t, irrespective of its location. We will use the Fokker–Planck approximation in the form given by Eq. (3.141) for the collision integral. The collision density $\sigma_s\Phi$ in the left-hand side of Eq. (5.61) cancels out, because it is also present in the Fokker–Planck formula for the collision integral.

This transforms the Boltzmann equation into a differential equation which has the form of the diffusion equation in two dimensions

$$\frac{\partial}{\partial t}\Phi\left(u_x,u_y,t\right) = \frac{1}{4}\sigma_s\langle\theta_0^2\rangle\left(\frac{\partial^2}{\partial u_x^2}+\frac{\partial^2}{\partial u_y^2}\right)\Phi\left(u_x,u_y,t\right). \tag{5.62}$$

It describes diffusion in the angular space (u_x, u_y), and instead of time in this equation we have distance t traveled by the particle. The diffusion coefficient is

$$D_{\text{diff.}} = \frac{1}{4}\sigma_s\langle\theta_0^2\rangle. \tag{5.63}$$

We consider a particle that started at point $t = 0$. At this point the angular distribution was a delta-function

$$\Phi\left(u_x, u_y, 0\right) = \delta\left(u_x\right)\delta\left(u_y\right). \tag{5.64}$$

The above equation is the initial condition for the diffusion equation. The boundary condition is that at infinity $(u_x \to \pm\infty, u_y \to \pm\infty)$ the fluence is zero. The diffusion equation and its solution are well known. With the above boundary conditions, the solution can be found by using a simple method based on two-dimensional Laplace transform, $\mathcal{L}_2$ (Abramowitz and Stegan 1964; Aghili and Moghaddam 2011). Because the range of each of the two variables u_x and u_y is $(-\infty, +\infty)$, we will use the bilateral (two-sided) form of the transform (LePage 2010). The properties of the two-dimensional bilateral Laplace transform are similar to those of the one-dimensional unilateral (one-sided) Laplace transform that we used earlier in this book, in Sect. 5.2.3. Furthermore, the method for solving the differential equation, Eq. (5.62), that we present below is very similar to the method we used in Sect. 5.2.2, Eqs.(5.9)–(5.16). The initial condition for the Laplace transform of the fluence $\Phi(p_x, p_y, 0)$ follows from Eq. (5.64)

$$\Phi\left(p_x, p_y, 0\right) \equiv \mathcal{L}_2\left\{\Phi\left(u_x, u_y, 0\right)\right\} = 1. \tag{5.65}$$

The Laplace transform of Eq. (5.62) is

$$\frac{\partial}{\partial t}\Phi\left(p_x, p_y, t\right) = \frac{1}{4}\sigma_s\langle\theta_0^2\rangle\left(p_x^2 + p_y^2\right)\Phi\left(p_x, p_y, t\right). \tag{5.66}$$

This equation with the initial condition given by Eq. (5.65) is easily integrated, giving a simple expression for the Laplace transform of the fluence

$$\Phi\left(p_x, p_y, t\right) = \exp\left[\frac{1}{4}\sigma_s\langle\theta_0^2\rangle t\left(p_x^2 + p_y^2\right)\right] \tag{5.67}$$

We can now find the fluence $\Phi(u_x, u_y, t)$ by performing the inverse transform $\mathcal{L}_2^{-1}$ (LePage 2010; Aghili and Moghaddam 2011):

$$\begin{aligned}\Phi\left(u_x, u_y, t\right) &= \mathcal{L}_2^{-1}\left\{\Phi(p_x, p_y, t)\right\} \\ &= \left(\frac{1}{2\pi i}\right)^2 \int_{c_x - i\infty}^{c_x + i\infty} \mathrm{d}p_x \int_{c_y - i\infty}^{c_y + i\infty} \mathrm{d}p_y \exp\left[\frac{1}{4}\sigma_s\langle\theta_0^2\rangle t\left(p_x^2 + p_y^2\right)\right].\end{aligned} \tag{5.68}$$

The above two-dimensional integral can obviously be rewritten as a product of two one-dimensional integrals over p_x and p_y. Each of the two integrals is calculated using the same technique that we used to calculate the integral in Eq. (5.11). The result is

$$\Phi\left(u_x, u_y, t\right) = \frac{1}{\pi\sigma_s\langle\theta_0^2\rangle t}\exp\left[-\frac{\left(u_x^2+u_y^2\right)}{\sigma_s\langle\theta_0^2\rangle t}\right]. \tag{5.69}$$

This is a two-dimensional normal distribution. Finally, for generating a particle trajectory, the angular variables (θ, ϕ) are more convenient than (u_x, u_y). In terms of θ and ϕ the solution is

$$\Phi\left(\theta, \phi, t\right) = \frac{1}{\pi\sigma_s\langle\theta_0^2\rangle t}\exp\left(-\frac{\theta^2}{\sigma_s\langle\theta_0^2\rangle t}\right). \tag{5.70}$$

The solution does not depend explicitly on the azimuthal angle ϕ. In other words it has an azimuthal symmetry. This form of the solution satisfies two obvious conditions. It is normalized as follows

$$\int_0^{2\pi} \mathrm{d}\phi \int_0^{\infty} \theta \mathrm{d}\theta\, \Phi\left(t, \theta\right) = 1, \tag{5.71}$$

and the width of the angular distribution, $\sqrt{\sigma_s\langle\theta_0^2\rangle t/2}$, tends to zero when $t \to 0$. We need to remind why in the above integral we have the factor θ. The factor appeared in the collision integral in Eq. (3.131) when we applied a small angle approximation to the integral over $\mu = \cos\theta$

$$\mathrm{d}\mu = -\sin\theta \mathrm{d}\theta \approx -\theta \mathrm{d}\theta. \tag{5.72}$$

We have also replaced the upper integration limit π with infinity because, in a small angle approximation, the number of particles scattered to large angles close to π can be assumed to be negligible, which is again consistent with Eq. (3.131).

We can now see that, in the Fokker–Planck approximation, the angular distribution is normal. However, this is a two-dimensional distribution. Hence, sampling the polar angle θ from a one-dimensional normal distribution would be an error. The correct interpretation is that angle θ is the length of a two-dimensional vector $\vec{u}$. The two components of the vector, u_x and u_y, are independent random variables, because the distribution in Eq. (5.69) can be written as a product of two one-dimensional distributions. Each component is normally distributed in one dimension. Methods for sampling the length of a vector were discussed in Chap. 2. The azimuthal angle ϕ is distributed uniformly in $[0, 2\pi)$.

Our result, obviously, simplifies the sampling of angular distributions in Monte Carlo algorithms. The only information about scattering cross sections needed to implement this algorithm is the product $\sigma_s\langle\theta_0^2\rangle$. As we noted in Sect. 3.9.6, [see Eq. (3.140)], this product is closely related to the transport cross section σ_{tr}.

5.3.2 *The Molière Theory*

The Molière model (Molière 1947; Molière 1948; Bethe 1953) is more accurate than the Fokker–Planck approximation and is relatively easy to implement in a Monte Carlo code. The angular distribution in this model is approximated by the sum of three universal functions that do not explicitly depend on the step size and can be tabulated in a form convenient for sampling the scattering angle. The limitations are the small scattering angle approximation and neglected energy losses. The latter set an upper limit on the step size. There is also a lower limit on the step size: the width of the angular distribution at the end of a step should be much greater than that following a single collision. Our derivation of the model is rather lengthy, as we explain every step and the underlying approximations in detail. It is based on applying the Fourier–Bessel transform to a small-angle version of the Boltzmann equation.

We begin with the Lagrangian form of the Boltzmann equation, Eq. (3.61), integrated over $\vec{r}$. Given that the fluence at infinity is zero, this integration eliminates the streaming operator. We assume an azimuthal symmetry and neglect energy losses. The equation then can be written in the following form:

$$\frac{\partial}{\partial t}\Phi\left(t,\mu\right)+\sigma\left(t\right)\Phi\left(t,\mu\right)=S\left(t,\mu\right)+\int \mathrm{d}\vec{\Omega}\,'\sigma_s\left(t,\vec{\Omega}\,'\cdot\vec{\Omega}\right)\Phi\left(t,\mu'\right). \tag{5.73}$$

Here t is the distance traveled by the particle. We also assume, for simplicity, that cross sections do not depend on t, and that scattering is the only interaction type, meaning that $\sigma_s=\sigma$. We then switch variables from $\mu=\cos\theta$ to θ and apply the small angle approximation, Eq. (5.72), to the collision integral:

$$\int \mathrm{d}\vec{\Omega}\,'\ldots=\int_0^{2\pi}\mathrm{d}\phi'\int_{-1}^{1}\mathrm{d}\mu'\ldots\approx\int_0^{2\pi}\mathrm{d}\phi'\int_0^{\infty}\theta'\mathrm{d}\theta'\ldots, \tag{5.74}$$

where we extended the upper integration limit in θ' from π to ∞, because in the small angle approximation, scattering to large angles close to π can be neglected. We further note that

$$\begin{aligned}\vec{\Omega}\,'\cdot\vec{\Omega}&=\cos\theta\cos\theta'+\sin\theta\sin\theta'\cos\left(\phi-\phi'\right)\\&\approx\left(1-\frac{\theta^2}{2}\right)\left(1-\frac{\theta'^2}{2}\right)+\theta\theta'\cos\left(\phi-\phi'\right)\\&\approx 1-\frac{\theta^2}{2}-\frac{\theta'^2}{2}+\theta\theta'\cos\left(\phi-\phi'\right),\end{aligned} \tag{5.75}$$

and, on the other hand,

$$\vec{\Omega}' \cdot \vec{\Omega} \equiv \mu_0 = \cos\theta_0 \approx 1 - \frac{\theta_0^2}{2}, \tag{5.76}$$

where θ_0 is the scattering angle. Combining Eqs. (5.75) and (5.76), we find that

$$\theta_0 = \sqrt{\theta^2 + \theta'^2 - 2\theta\theta' \cos(\phi - \phi')}. \tag{5.77}$$

As for the source function, we choose a monodirectional point source, parallel to the z axis:

$$S(t, \theta) = \frac{1}{2\pi\theta} \delta(t)\, \delta(\theta). \tag{5.78}$$

It is normalized as follows:

$$\int_0^\infty \mathrm{d}t \int_0^{2\pi} \mathrm{d}\phi \int_0^\infty \theta \mathrm{d}\theta\, S(t, \theta) = 1. \tag{5.79}$$

This transforms Eq. (5.73) into the Boltzmann equation in the small-angle approximation. For $t > 0$, where the source function is zero, we have

$$\frac{\partial}{\partial t}\Phi(t, \theta) + \sigma_s \Phi(t, \theta) = \int_0^{2\pi} \mathrm{d}\phi' \int_0^\infty \theta' \mathrm{d}\theta' \sigma_s(\theta_0)\, \Phi\left(t, \theta'\right). \tag{5.80}$$

This is yet another integro-differential equation. We have previously used several solution methods. Here we introduce one more, the Fourier–Bessel transform. The reason for choosing this transform is the weighting factor θ' in the collision integral in Eq. (5.80). A product of two Bessel functions J_0 integrated with this weighting satisfies the so-called closure equation (Arfken et al. 1985):

$$\int_0^\infty \theta \mathrm{d}\theta\, J_0(\eta\theta)\, J_0\left(\eta'\theta\right) = \frac{1}{\eta'} \delta\left(\eta - \eta'\right). \tag{5.81}$$

The closure equation plays the role of the orthogonality condition in other transforms, such as the Legendre transform. The Fourier–Bessel transform [or the Hankel transform (Poularikas 2010)] of the fluence according to the transform definition is as follows:

$$\Phi(t, \eta) = \int_0^\infty \theta \mathrm{d}\theta\, J_0(\eta\theta)\, \Phi(t, \theta). \tag{5.82}$$

It follows from Eqs. (5.81) and (5.82) that:

$$\Phi(t,\theta) = \int_0^\infty \eta \mathrm{d}\eta\, J_0(\eta\theta)\, \Phi(t,\eta)\,, \tag{5.83}$$

which defines the inverse transform.

We now apply the Fourier–Bessel transform to Eq. (5.80). The transform of the left-hand side is elementary. That of the collision integral involves several steps, but the final result is simple. The transform of the collision integral is given by the following integral:

$$\int_0^\infty \theta \mathrm{d}\theta\, J_0(\eta\theta) \int_0^{2\pi} \mathrm{d}\phi' \int_0^\infty \theta' \mathrm{d}\theta' \sigma_s(\theta_0)\, \Phi\left(t,\theta'\right). \tag{5.84}$$

To simplify, we replace the differential scattering cross section $\sigma_s(\theta_0)$ with its inverse transform representation:

$$\sigma_s(\theta_0) = \int_0^\infty \eta' \mathrm{d}\eta'\, J_0\left(\eta'\theta_0\right) \sigma_s\left(\eta'\right), \tag{5.85}$$

and then use the addition theorem (Macdonald 1900):

$$J_0\left(\eta'\theta_0\right) = J_0\left(\eta'\theta\right) J_0\left(\eta'\theta'\right) + 2\sum_{k=1}^{\infty} J_k\left(\eta'\theta\right) J_k\left(\eta'\theta'\right) \cos\left[k(\phi-\phi')\right], \tag{5.86}$$

which is valid for θ_0 defined by Eq. (5.77). When we integrate the above sum over ϕ from 0 to 2π, the integral of the cosine is zero for any $k \geq 1$, and therefore we can leave out the sum. Next, we insert Eqs. (5.85) and (5.86) into the collision integral, Eq. (5.84):

$$\int_0^\infty \theta \mathrm{d}\theta\, J_0(\eta\theta) \int_0^{2\pi} \mathrm{d}\phi' \int_0^\infty \theta' \mathrm{d}\theta' \times \left[\int_0^\infty \eta' \mathrm{d}\eta'\, J_0\left(\eta'\theta\right) J_0\left(\eta'\theta'\right) \sigma_s\left(\eta'\right)\right] \Phi\left(t,\theta'\right) \tag{5.87}$$

The integral over θ produces a delta function, $\delta(\eta-\eta')/\eta'$, because of the closure equation. This eliminates the integral over η'. Finally, the remaining integral, over θ', is nothing but the transform $\Phi(t,\eta)$ of the fluence. Together with the transformed left-hand side of Eq. (5.80) this produces a simple differential equation for the fluence transform:

$$\frac{\partial}{\partial t}\Phi(t,\eta) + \sigma_s \Phi(t,\eta) = 2\pi\sigma_s(\eta)\, \Phi(t,\eta)\,. \tag{5.88}$$

Given that the total cross section σ_s is an integral of the differential cross section over ϕ and θ, we can write the solution of the above equation as follows:

$$\Phi(t,\eta) = \frac{1}{2\pi}\exp\left\{-2\pi t\int_0^\infty \theta \mathrm{d}\theta\,[1 - J_0(\eta\theta)]\,\sigma_s(\theta)\right\}. \tag{5.89}$$

The multiplier $(1/2\pi)$ in front of the exponent in Eq. (5.89) was chosen so that the solution in the limit $t \to 0$ was consistent with the Fourier–Bessel transform of the source function, Eq. (5.78).

The two remaining steps are choosing an appropriate scattering cross section and performing the inverse Fourier–Bessel transform. The cross section in this model is given by the Rutherford formula corrected for screening of the nucleus by the atomic electrons. Examples of screening corrections can be found, for example, in Nikjoo et al. (2008). In this formalism, however, we do not need to choose a specific form of the screening correction, because ultimately screening is characterized by a single parameter. Hence, we write the cross section in the following general form:

$$\sigma_s(\theta) = A\frac{q(\theta)}{\theta^4}, \tag{5.90}$$

where $q(\theta)$ accounts for screening. Screening is important only for relatively large impact parameters, corresponding to small scattering angles comparable to (Bethe, 1953):

$$\chi_0 = \frac{\lambda/2\pi}{0.885\,a_0\,Z^{-1/3}}, \tag{5.91}$$

where λ is the de Broglie wavelength of the incident particle, a_0 the Bohr radius, and Z the atomic number of the atom. The expression in the denominator is the Fermi radius of the atom. For angles much greater than χ_0, screening is negligible, and $q(\theta) = 1$.

Before we can perform the inverse transform, we simplify the integral in the exponent, Eq. (5.89). First, we split it into two integrals, one from an angle k to infinity and the other from zero, to k. The angle k is chosen so that, on one hand, it is much greater than χ_0 (in which case, $q(\theta) = 1$ in the first integral), and on the other hand, it is sufficiently small to ensure that $\eta\theta \ll 1$ in the second integral. We will show that the latter condition is satisfied when $k \ll \theta_c$, where the parameter θ_c characterizes the width of the angular distribution $\Phi(t,\theta)$. Then, angle k can satisfy both requirements only when $\chi_0 \ll \theta_c$. Given that χ_0 characterizes a single scattering event, the latter condition requires that the average number of collisions the particle undergoes as it travels a distance t is much greater than one. The number of collisions, however, cannot be too large, because we neglect particle energy losses. The latter condition sets an upper bound on the step size t.

We now show that $\eta\theta \ll 1$ is satisfied in the second integral (integral over θ from 0 to k) when $k \ll \theta_c$. First, we note that $\eta\theta \ll 1$ is satisfied when $\eta k \ll 1$.

Second, we note that η can in principle be any positive number. Then, following the terminology of Bethe (1953) we consider only "important values" of η, that is, only those values that contribute significantly to the integral in Eq. (5.83) (the inverse transform). These are largely determined by the inverse fluence transform $\Phi(t, \eta)$. To estimate what "important values" are, we approximate the angular distribution by a normal distribution:

$$\Phi(t, \theta) = \exp\left[-\left(\frac{\theta}{\theta_c}\right)^2\right], \tag{5.92}$$

where $\theta_c = \theta_c(t)$ is a parameter characterizing the width of the distribution. For this distribution we can find the Fourier–Bessel transform analytically:

$$\Phi(t, \eta) = \int_0^\infty \theta \mathrm{d}\theta \, J_0(\eta\theta) \exp\left[-\left(\frac{\theta}{\theta_c}\right)^2\right]. \tag{5.93}$$

The substitution $w = \eta\theta$ transforms this integral into a known integral (Gradshteyn and Ryzhik 1980):

$$\int_0^\infty w \mathrm{d}w \, \mathrm{e}^{-aw^2} J_0(w) = \frac{1}{2a} \exp\left(-\frac{1}{4a}\right), \tag{5.94}$$

leading to:

$$\Phi(t, \eta) = \frac{\theta_c^2}{2} \exp\left[-\left(\frac{\eta\theta_c}{2}\right)^2\right]. \tag{5.95}$$

This result shows that important values of η are of the order of $1/\theta_c$ or less. Then, in $\eta k \ll 1$ we can replace η by approximately the largest of its important values $\eta = 1/\theta_c$, which brings us to the condition $k \ll \theta_c$. With this condition satisfied, we can assume that the argument of the Bessel function in the integral from 0 to k is small, and approximate it using the first two terms of the Taylor series:

$$\int_0^k \theta \mathrm{d}\theta \, [1 - J_0(\eta\theta)] \frac{q(\theta)}{\theta^4} \approx \int_0^k \theta \mathrm{d}\theta \frac{1}{4}\eta^2\theta^2 \frac{q(\theta)}{\theta^4} = \frac{\eta^2}{4} \int_0^k \mathrm{d}\theta \frac{q(\theta)}{\theta} \equiv \frac{\eta^2}{4} I_2(k), \tag{5.96}$$

where we introduced an obvious notation $I_2(k)$ for the integral. In the above equation and further in this section we use identities involving Bessel functions J_0 and J_1 that can be found in Abramowitz and Stegan (1964).

A property that we will use later is that for sufficiently large k ($k \gg \chi_0$), the difference $I_2(k) - \ln k$ is independent of k:

$$\frac{\partial}{\partial k}\left[\int_0^k \frac{q(\theta)}{\theta} \mathrm{d}\theta - \ln k\right] = \frac{q(k)}{k} - \frac{1}{k} \to 0. \tag{5.97}$$

In the integral from k to infinity we neglect screening entirely, and set $q(\theta) = 1$. The following algebra may seem complicated, but is in fact a sequence of three integrations by parts, yielding the known integral given by Eq. (5.94). First, we change the integration variable from θ to $w = \eta\theta$:

$$\int_k^\infty \theta \mathrm{d}\theta \left[1 - J_0(\eta\theta)\right] \frac{1}{\theta^4} = \eta^2 \int_{k\eta}^\infty \frac{\mathrm{d}w}{w^3} \left[1 - J_0(w)\right]. \tag{5.98}$$

We then do the first integration by parts:

$$\int_{k\eta}^\infty \frac{\mathrm{d}w}{w^3} \left[1 - J_0(w)\right] = \frac{1 - J_0(k\eta)}{2k^2\eta^2} + \frac{1}{2} \int_{k\eta}^\infty \frac{\mathrm{d}w}{w^2} J_1(w), \tag{5.99}$$

where we used $\mathrm{d}J_0(w)/\mathrm{d}w = -J_1(w)$. The second integration by parts follows:

$$\frac{1}{2} \int_{k\eta}^\infty \frac{\mathrm{d}w}{w^2} J_1(w) = \frac{J_1(k\eta)}{4k\eta} + \frac{1}{4} \int_{k\eta}^\infty \frac{\mathrm{d}w}{w} J_0(w). \tag{5.100}$$

where we used $\mathrm{d}J_1(w)/\mathrm{d}w = J_0(w) - J_1(w)/w$. The final integration then gives

$$\frac{1}{4} \int_{k\eta}^\infty \frac{\mathrm{d}w}{w} J_0(w) = -\frac{1}{4} \ln(k\eta) J_0(k\eta) + \frac{1}{4} \int_{k\eta}^\infty \mathrm{d}w \ln(w) J_1(w). \tag{5.101}$$

Next, we insert Eqs. (5.99)–(5.101) back into Eq. (5.98) and write the result assuming that $k\eta$ is small. This is because we are still concerned only with important η values. More specifically, we retain only the terms that are nonvanishing in the limit $k\eta \to 0$. We can then use:

$$\int_{k\eta}^\infty \mathrm{d}w \ln(w) J_1(w) \approx \int_0^\infty \mathrm{d}w \ln(w) J_1(w) = \ln 2 - \gamma, \tag{5.102}$$

where γ is the Euler constant. This integral can be found in Gradshteyn and Ryzhik (1980). Also, for small w we use $J_1(w) \approx w/2$. Then, the final result for the integral is:

$$\int_{k\eta}^\infty \frac{\mathrm{d}w}{w^3} \left[1 - J_0(w)\right] \approx \frac{1}{4} \left[1 - \ln(k\eta) + \ln 2 - \gamma\right]. \tag{5.103}$$

To find the transform of the fluence, we insert Eqs. (5.96) and (5.103) back into Eq. (5.89), which gives

$$\Phi(t, \eta) = \frac{1}{2\pi} \exp\left\{-\frac{\pi}{2} A t \eta^2 \left[I_2(k) + 1 - \ln(k\eta) + \ln 2 - \gamma\right]\right\}. \tag{5.104}$$

Before performing the inverse transform, we introduce some notations, starting with

$$\chi_c^2 = \pi A t. \tag{5.105}$$

It will be shown below that χ_c^2 characterizes the width of the angular distribution of the fluence. The next quantity, χ_a, is defined so as to be k-independent for sufficiently large k by virtue of Eq. (5.97):

$$-\ln \chi_a = I_2(k) - \ln k + \frac{1}{2}. \tag{5.106}$$

The last three quantities, b, B, and u, are introduced only for convenience and are defined as follows:

$$b = \ln\left(\frac{\chi_c^2}{\chi_a^2}\right) + 1 - 2\gamma. \tag{5.107}$$

$$b = B - \ln B. \tag{5.108}$$

$$u = \chi_c \eta \sqrt{B}. \tag{5.109}$$

In this notation, Eq. (5.104) takes a simple form:

$$\Phi(t, \eta) = \frac{1}{2\pi} \exp\left[-\frac{u^2}{4} + \frac{u^2}{4B} \ln\left(\frac{u^2}{4}\right)\right]. \tag{5.110}$$

Finally, to find the fluence, we need to perform the inverse transform, Eq. (5.83):

$$\Phi(t, \theta) = \frac{1}{2\pi\chi_c^2 B} \int_0^\infty u\mathrm{d}u\, J_0\left(\frac{\theta}{\chi_c\sqrt{B}}u\right) \exp\left[-\frac{u^2}{4} + \frac{u^2}{4B} \ln\left(\frac{u^2}{4}\right)\right], \tag{5.111}$$

where we switched the integration variable from η to u. This integral can be calculated numerically and tabulated as a function of two parameters: B and $\theta/\chi_c\sqrt{B}$. This table can be in the form of the cumulative distribution, making the sampling of the scattering angle θ straightforward. Practically, however, a different method that has been used is based on the expansion of the second half of the exponent into a series:

$$\Phi(t, \theta) = \frac{1}{2\pi\chi_c^2 B} \sum_{n=0}^{\infty} \frac{1}{B^n} f^{(n)}\left(\frac{\theta}{\chi_c\sqrt{B}}\right), \tag{5.112}$$

where

$$f^{(n)}(\vartheta) = \frac{1}{n!} \int_0^\infty u\mathrm{d}u\, J_0(\vartheta u) \exp\left(-\frac{u^2}{4}\right) \left[\frac{u^2}{4} \ln\left(\frac{u^2}{4}\right)\right]^n. \tag{5.113}$$

Consider the first term in the sum, $n = 0$:

$$\Phi_0(t,\theta) = \frac{1}{2\pi\chi_c^2 B}\int_0^\infty u\mathrm{d}u\, J_0\left(\frac{\theta}{\chi_c\sqrt{B}}u\right)\exp\left(-\frac{u^2}{4}\right). \tag{5.114}$$

Through the substitution $w = \theta u/\chi_c B$ the above integral takes a form that can be integrated analytically, Eq. (5.94), leading, yet again, to the normal distribution:

$$\Phi_0(t,\theta) = \frac{1}{\pi B\chi_c^2}\exp\left(-\frac{\theta^2}{B\chi_c^2}\right). \tag{5.115}$$

Higher order functions $f^{(n)}$, with $n = 1, 2\ldots$, are not as simple. Fortunately, the first three terms of the sum in Eq. (5.112), with $n = 0, 1, 2$, yield an accuracy of the angular distribution of about 1 % or better for any angle, and tables of $f^{(0)}, f^{(1)}$, and $f^{(2)}$ are available in the literature (Bethe 1953).

5.3.3 *The Goudsmit–Saunderson Distribution*

This method was introduced by Goudsmit and Saunderson (1940a, 1940b). In contrast to the Fokker–Planck and the Molière models, it does not rely on a small angle approximation. This method yields more accurate angular distributions, albeit usually at the expense of increased computing time. For simplicity, we will derive it in a form that neglects energy losses.

The method is relatively straightforward on the whole. First, the angular dependence of the scattering cross section is represented as a Legendre series. Then, the angular distribution after multiple scatterings is found analytically, as an infinite sum. This solution is exact, in principle. However, in a numerical implementation, only a finite number of terms can be retained in the sum.

We start with the same form of the Boltzmann equation as in the preceding section, Eq. (5.73). We neglect energy losses and assume that the medium is homogeneous and scattering is the only process, that is $\sigma(z) = \sigma_s$. This time, however, we do not assume small scattering angles. The Boltzmann equation then takes the following form.

$$\frac{\partial}{\partial t}\Phi(t,\mu) + \sigma_s\Phi(t,\mu) = S(t,\mu) + \int_0^{2\pi}\mathrm{d}\phi'\int_{-1}^{1}\mathrm{d}\mu'\sigma_s(\mu_0)\,\Phi\left(t,\mu'\right), \tag{5.116}$$

where $\mu_0 = (\vec{\Omega}'\cdot\vec{\Omega})$ is the cosine of the scattering angle. The source is located at $t = 0$ and emits particles in the positive z direction, $\cos\theta = \mu = 1$:

$$S(t,\mu) = \frac{1}{2\pi}\delta(t)\,\delta(\mu - 1). \tag{5.117}$$

The collision integral is simplified using a Legendre series expansion of the angular dependence of the scattering cross section $\sigma_s(\mu_0)$. The algebra is exactly the same as in Sect. 3.9.7, and the result is given by Eq. (3.152). Then, for $t > 0$, the Boltzmann equation can be written in the following form, which can be solved analytically:

$$\frac{\partial}{\partial t}\Phi(t,\mu) + \sigma_s \Phi(t,\mu) = \sum_{n=0}^{\infty} \frac{4\pi}{2n+1} P_n(\mu)\,\sigma_n \Phi_n(t)\,. \tag{5.118}$$

To find the solution, in a similar fashion to Sect. 3.9.7, we multiply this equation by $P_k(\mu)(2k+1)/2$ and integrate over μ from -1 to 1. The left-hand side produces Legendre coefficients of the fluence, Φ_k, and the collision integral is further simplified using the orthogonality of Legendre polynomials, Eq. (3.151). The result is a simple differential equation for Φ_k:

$$\frac{\partial}{\partial t}\Phi_k(t) + \sigma_s \Phi_k(t) = \frac{4\pi}{2k+1}\sigma_k \Phi_k(t)\,, \quad k = 0, 1, 2\ldots \tag{5.119}$$

The solution of this equation is the exponent:

$$\Phi_k(t) = A_k \exp(-\tilde{\sigma}_k t), \tag{5.120}$$

where

$$\tilde{\sigma}_k = \sigma_s - \frac{4\pi}{2k+1}\sigma_k = 2\pi \int_{-1}^{1} \mathrm{d}\mu\, \sigma_s(\mu)\,[1 - P_k(\mu)]\,. \tag{5.121}$$

In the literature, $\tilde{\sigma}_k$ is often replaced by a closely related quantity $Q_k = \tilde{\sigma}_k/\sigma$. To find the integration constants A_k, we note that when the distance t approaches zero from the right, the angular distribution $\Phi(t,\mu)$ must approach that of the source function $S(t,\mu)$. The Legendre coefficients behave in the same way, that is:

$$A_k = \lim_{t\to 0+} \Phi_k(t) = \frac{2k+1}{2} \int_{-1}^{1} \mathrm{d}\mu\, \frac{1}{2\pi}\delta(\mu - 1) P_k(\mu) = \frac{2k+1}{4\pi}. \tag{5.122}$$

If the Legendre coefficients $\tilde{\sigma}_k$ are known, the fluence can be calculated by performing the inverse Fourier–Legendre transform:

$$\Phi(t,\mu) = \sum_{k=0}^{\infty} \Phi_k(t) P_k(\mu) = \sum_{k=0}^{\infty} \frac{2k+1}{4\pi} \exp(-\tilde{\sigma}_k t) P_k(\mu)\,. \tag{5.123}$$

As can be easily verified, this solution remains normalized for any $t \geq 0$:

$$\int_0^{2\pi} \mathrm{d}\phi \int_{-1}^{1} \mathrm{d}\mu\, \Phi(t,\mu) = 1. \tag{5.124}$$

To derive this result we multiplied Eq. (5.123) by $P_0 = 1$, then performed the integration where we used the orthogonality property of the Legendre polynomials. The integration result included $\tilde{\sigma}_0$, which is zero according to its definition in Eq. (5.121). The fluence remains normalized because scattering was the only process included in the model.

Moments of the angular distribution $\langle \mu^k(t) \rangle$ ($k = 1, 2, \ldots$) are easily calculated using the orthogonality property of the Legendre polynomials, for example

$$\begin{aligned} \langle \mu\,(t) \rangle &= \langle P_1\,(\mu) \rangle = \int_{-1}^{1} \mathrm{d}\phi \int_{-1}^{1} \mathrm{d}\mu\ \Phi\,(t,\mu)\, P_1\,(\mu) \\ &= \sum_{k=0}^{\infty} \frac{2k+1}{2} \exp\left(-\tilde{\sigma}_k t\right) \int_{-1}^{1} \mathrm{d}\mu\ P_k\,(\mu)\, P_1\,(\mu) = \exp\left(-\tilde{\sigma}_1 t\right). \end{aligned} \tag{5.125}$$

Similarly,

$$\langle \mu^2\,(t) \rangle = \left\langle \frac{2}{3} P_2\,(\mu) + \frac{1}{3} \right\rangle = \frac{2}{3} \exp\left(-\tilde{\sigma}_2 t\right) + \frac{1}{3}. \tag{5.126}$$

The Goudsmit–Saunderson solution was found by assuming that the step size t is so small that any changes in $\tilde{\sigma}_k$ can be neglected. If these changes need to be accounted for, then the differential equation for the Legendre coefficients of the fluence $\Phi_k(t)$, Eq. (5.119), can be solved without assuming constant cross sections. In that case $\tilde{\sigma}_k$ is a function of the distance traveled by the particle, $\tilde{\sigma}_k = \tilde{\sigma}_k(t)$, and in the solution the product $\tilde{\sigma}_k t$ in the exponent is replaced by an integral of $\tilde{\sigma}_k(t)$ over the length of the step, that is from 0 to t. This integration can account, for example, for material inhomogeneity and/or energy losses. In the latter case the continuous slowing down approximation can be used to determine particle energy as a function of the distance it has traveled.

Only in a few special cases (e.g., isotropic scattering) can the sum in Eq. (5.123) be calculated analytically. In numerical calculations the sum must be terminated at a finite $k = k_{\max}$. To achieve good accuracy when scattering is strongly forward-peaked, which is the case for charged particles, a large $k_{\max}$ is required, $k_{\max} \gg 1$.

5.4 Spatial Distribution

In this section we calculate the first two moments of the spatial distribution of the fluence for a charged particle that has traveled distance t. The methodology that we use was developed by Lewis (1950). An important application of these results is in the evaluation of the spatial accuracy of condensed history algorithms. In such algorithms, the continuous trajectory of a charged particle is approximated by steps of finite size. Choosing the step size is an important aspect of algorithm optimization. From the results presented in this section, uncertainties in the spatial

distribution can be readily estimated as functions of the step size. Uncertainties in the angular and energy distributions should also be evaluated using one of the multiple scattering theories presented in this chapter.

5.4.1 Longitudinal Displacement

We start with a simplified version of the Lagrangian form of the transport equation, Eq. (3.61). For brevity we leave out the source function, the initial phase coordinates of the particle, $\vec{r}_0$, $\vec{\Omega}_0$, E_0, and its energy E. We also assume that the material is homogeneous, and that $\sigma = \sigma_s$:

$$\frac{\partial}{\partial t}\Phi\left(\vec{r},\vec{\Omega},t\right) + \left(\vec{\Omega}\cdot\vec{\nabla}\right)\Phi\left(\vec{r},\vec{\Omega},t\right) + \sigma_s\Phi\left(\vec{r},\vec{\Omega},t\right) = \int \mathrm{d}\vec{\Omega}\,'\sigma_s\left(\vec{\Omega}\cdot\vec{\Omega}\,'\right)\Phi\left(\vec{r},\vec{\Omega}\,',t\right). \tag{5.127}$$

For simplicity we also neglect energy losses. However, in this formalism, energy losses can be accounted for using the continuous slowing down approximation. This is explained later in this section.

We seek a solution in the following form [some authors call it the Laplace series (Arfken et al. 2013)]:

$$\Phi\left(\vec{r},\vec{\Omega},t\right) = \sum_{l=0}^{\infty}\sum_{m=-l}^{l}\Phi_{l,m}\left(\vec{r},t\right)Y_{l,m}\left(\vec{\Omega}\right), \tag{5.128}$$

where

$$\Phi_{l,m}\left(\vec{r},t\right) = \int \mathrm{d}\vec{\Omega}\,Y^*_{l,m}\left(\vec{\Omega}\right)\Phi\left(\vec{r},\vec{\Omega},t\right), \tag{5.129}$$

and $Y_{l,m}$ and $Y^*_{l,m}$ are the spherical harmonic and its complex conjugate. To derive the equation for the coefficients $\Phi_{l,m}(\vec{r},t)$ we multiply Eq. (5.127) by $Y^*_{l,m}(\vec{\Omega})$ and integrate over $\vec{\Omega}$. The integration of the first and last terms on the left-hand side of the equation is elementary. The streaming operator is integrated as follows:

$$\begin{aligned}\int \mathrm{d}\vec{\Omega}\,Y^*_{l,m}\left(\vec{\Omega}\right)\left(\vec{\Omega}\cdot\vec{\nabla}\right)\Phi\left(\vec{r},\vec{\Omega},t\right) &= \int \mathrm{d}\vec{\Omega}\,Y^*_{l,m}\left(\vec{\Omega}\right)\left(\vec{\Omega}\cdot\vec{\nabla}\right)\sum_{l'=0}^{\infty}\sum_{m'=-l'}^{l'}\Phi_{l',m'}\left(\vec{r},t\right)Y_{l',m'}\left(\vec{\Omega}\right)\\ &= \sum_{l'=0}^{\infty}\sum_{m'=-l'}^{l'}\left[\vec{\nabla}\Phi_{l',m'}\left(\vec{r},t\right)\right]\cdot\int \mathrm{d}\vec{\Omega}\,Y^*_{l,m}\left(\vec{\Omega}\right)\vec{\Omega}\,Y_{l',m'}\left(\vec{\Omega}\right)\\ &= \sum_{l',m'}\left[\vec{\nabla}\Phi_{l',m'}\left(\vec{r},t\right)\right]\cdot\vec{Q}_{l,m,l',m'},\end{aligned} \tag{5.130}$$

where we introduced the notation:

$$\vec{Q}_{l,m,l',m'} \equiv \int \mathrm{d}\vec{\Omega}\; Y^*_{l,m}\left(\vec{\Omega}\right)\; \vec{\Omega}\; Y_{l',m'}\left(\vec{\Omega}\right). \tag{5.131}$$

As for the collision integral in Eq. (5.127), we write the Legendre series for the scattering cross section, σ_s:

$$\sigma_s\left(\vec{\Omega}\cdot\vec{\Omega}\,'\right) = \sum_{n=0}^{\infty}\sigma_n P_n\left(\vec{\Omega}\cdot\vec{\Omega}\,'\right), \tag{5.132}$$

and then apply the addition theorem (Arfken et al. 2013)

$$P_n\left(\vec{\Omega}\cdot\vec{\Omega}\,'\right) = \frac{4\pi}{2n+1}\sum_{k=-n}^{n} Y_{n,k}\left(\vec{\Omega}\right) Y^*_{n,k}\left(\vec{\Omega}\,'\right) \tag{5.133}$$

to replace the Legendre polynomials P_n with spherical harmonics. Then we insert Eqs. (5.132) and (5.133) into the collision integral, multiply the integral by $Y^*_{l,m}(\vec{\Omega})$, and integrate over $\vec{\Omega}$:

$$\int \mathrm{d}\vec{\Omega}\; Y^*_{l,m}\left(\vec{\Omega}\right)\int \mathrm{d}\vec{\Omega}'\sum_{n=0}^{\infty}\sigma_n\frac{4\pi}{2n+1}\sum_{k=-n}^{n} Y_{n,k}\left(\vec{\Omega}\right) Y^*_{n,k}\left(\vec{\Omega}'\right)\Phi\left(\vec{r},\vec{\Omega}',t\right). \tag{5.134}$$

When we integrate over $\vec{\Omega}$ we can use the orthogonality property of spherical harmonics:

$$\int \mathrm{d}\vec{\Omega}\; Y^*_{l',m'}\left(\vec{\Omega}\right) Y_{l,m}\left(\vec{\Omega}\right) = \delta_{l,l'}\delta_{m,m'}. \tag{5.135}$$

This eliminates the two sums. The remaining integral, over $\vec{\Omega}\,'$, is simply $\Phi_{l,m}(\vec{r},t)$. With this result for the collision integral and with the formula for the streaming operator, Eq. (5.130), the equation for the spherical coefficients of the fluence $\Phi_{l,m}$ takes the form:

$$\frac{\partial}{\partial t}\Phi_{l,m}(\vec{r},t) + \sum_{l',m'}\left[\vec{\nabla}\Phi_{l',m'}\left(\vec{r},t\right)\right]\cdot\vec{Q}_{l,m,l',m'} + \sigma_s\Phi_{l,m}(\vec{r},t) = \frac{4\pi}{2l+1}\sigma_l\Phi_{l,m}(\vec{r},t). \tag{5.136}$$

Finally, we combine the expression on the right-hand side with $\sigma_s\Phi_{l,m}$ using $\tilde{\sigma}_l$ defined in Eq. (5.121):

$$\frac{\partial}{\partial t}\Phi_{l,m}(\vec{r},t) + \sum_{l',m'}\left[\vec{\nabla}\Phi_{l',m'}\left(\vec{r},t\right)\right]\cdot\vec{Q}_{l,m,l',m'} + \tilde{\sigma}_l\Phi_{l,m}(\vec{r},t) = 0. \tag{5.137}$$

Equation (5.137) is the main equation of the formalism. Next, we consider the longitudinal and transverse distributions separately.

To derive the equation for the longitudinal distribution, we integrate Eq. (5.137) over x and y. This integration eliminates the derivatives $\partial/\partial x$ and $\partial/\partial y$ in the square brackets, as the fluence is zero for infinite x and y. This integration yields the following equation:

$$\frac{\partial}{\partial t}\Phi_{l,m}(z,t)+\sum_{l',m'}\left[\frac{\partial}{\partial z}\Phi_{l',m'}(z,t)\right]\left(Q_{l,m,l',m'}\right)_z+\tilde{\sigma}_l\Phi_{l,m}(z,t)=0. \tag{5.138}$$

Given that the direction of the z axis coincides with the initial particle direction, $\vec{\Omega}_0$, we can assume azimuthal symmetry. In this case, the fluence Φ is a function of the cosine μ of the polar angle, and does not depend on the azimuthal angle ϕ. We then need to find the solution $\Phi_{l,m}$ only for $m=0$. For all $m\neq 0$, the coefficients $\Phi_{l,m}$ are zero:

$$\begin{aligned}\Phi_{l,m}(z,t)&=\int_{-1}^{1}\mathrm{d}\mu\int_0^{2\pi}\mathrm{d}\phi\;\; Y^*_{l,m}(\mu,\phi)\,\Phi(z,\mu,t)\\&=\int_{-1}^{1}\mathrm{d}\mu\int_0^{2\pi}\mathrm{d}\phi\;\;\sqrt{\frac{2l+1}{4\pi}\frac{(l-m)!}{(l+m)!}}P_l^m(\mu)\,\mathrm{e}^{-im\phi}\,\Phi(z,\mu,t)\\&=0,\;\;\text{if}\;\;m\neq 0.\end{aligned} \tag{5.139}$$

Similarly, the matrix $\left(Q_{l,m,l',m'}\right)_z$ is simplified, because for $m=0$ we have

$$\begin{aligned}\left(Q_{l,0,l',m'}\right)_z=\int\mathrm{d}\vec{\Omega}\;Y^*_{l,0}\left(\vec{\Omega}\right)\Omega_z Y_{l',m'}\left(\vec{\Omega}\right)=\sqrt{\frac{2l+1}{4\pi}}\sqrt{\frac{2l'+1}{4\pi}\frac{(l'-m')!}{(l'+m')!}}\\\times\int_0^{2\pi}\mathrm{d}\phi\,\mathrm{e}^{im'\phi}\int_{-1}^{1}\mathrm{d}\mu\,P_l^0(\mu)\,\mu\,P_{l'}^{m'}(\mu)=0,\;\;\text{if}\;\;m'\neq 0,\end{aligned} \tag{5.140}$$

where we used $\Omega_z=\mu$. We thus only need to find

$$\left(Q_{l,0,l',0}\right)_z=\sqrt{\frac{2l+1}{4\pi}}\sqrt{\frac{2l'+1}{4\pi}}\int_0^{2\pi}\mathrm{d}\phi\int_{-1}^{1}\mathrm{d}\mu\,P_l(\mu)\,\mu\,P_{l'}(\mu)\,. \tag{5.141}$$

For any $l\geq 1$ this integral has only two nonzero values, when $l'=l-1$, and $l'=l+1$, see Eqs. (3.154)–(3.156):

$$\left(Q_{l,0,l-1,0}\right)_z=\frac{l}{\sqrt{4l^2-1}}\equiv\alpha_l; \tag{5.142}$$

$$\left(Q_{l,0,l+1,0}\right)_z=\frac{l+1}{\sqrt{4\,(l+1)^2-1}}\equiv\alpha_{l+1}. \tag{5.143}$$

For $l = 0$, there is only one nonzero value, when $l' = 1$:

$$(Q_{0,0,1,0})_z = \frac{1}{\sqrt{3}} \equiv \alpha_0. \tag{5.144}$$

With these results we can now write Eq. (5.138) for $m = 0$ and $l \geq 1$:

$$\frac{\partial}{\partial t}\Phi_{l,0}(z,t) + \frac{\partial}{\partial z}\left[\Phi_{l-1,0}(z,t)\,\alpha_l + \Phi_{l+1,0}(z,t)\,\alpha_{l+1}\right] + \tilde{\sigma}_l \Phi_{l,0}(z,t) = 0. \tag{5.145}$$

The equation for $l = 0$ is the same, apart from the first term in the square brackets being zero, and $\tilde{\sigma}_0$ is also zero. To solve Eq. (5.145), we apply boundary conditions:

$$\lim_{t\to 0+} \Phi(z,\mu,t) = \frac{\delta(z)}{2\pi}\delta(\mu - 1), \tag{5.146}$$

expressing that the facts that the particle started in the direction of z-axis ($\mu = 1$), and that at $t = 0$ the particle z-coordinate was zero, and the fluence was normalized:

$$\int_{-\infty}^{+\infty} \mathrm{d}z \int_{-1}^{1} \mathrm{d}\mu \int_{0}^{2\pi} \mathrm{d}\phi\ \Phi(z,\mu,0) = 1. \tag{5.147}$$

The boundary conditions for the coefficients $\Phi_{l,0}$ follow from the above boundary conditions for the fluence and from Eq. (5.129):

$$\lim_{t\to 0+} \Phi_{l,0}(z,t) = \frac{\delta(z)}{2\pi} \int_{-1}^{1} \mathrm{d}\mu \int_{0}^{2\pi} \mathrm{d}\phi\ \ Y^*_{l,0}(\mu,\phi)\,\delta(\mu - 1) = \delta(z)\sqrt{\frac{2l+1}{4\pi}}, \tag{5.148}$$

where we used $P_l(1) = 1$. The moments of the longitudinal distribution are by definition:

$$\langle z^n(t)\rangle = \frac{\int_{-\infty}^{+\infty} \mathrm{d}z\, z^n \int_0^{2\pi} \mathrm{d}\phi\ \int_{-1}^{1} \mathrm{d}\mu\ \Phi(z,\mu,t)}{\int_{-\infty}^{+\infty} \mathrm{d}z \int_0^{2\pi} \mathrm{d}\phi\ \int_{-1}^{1} \mathrm{d}\mu\ \Phi(z,\mu,t)};\ \ n = 1,2,\ldots \tag{5.149}$$

Comparing the above integrals with

$$\Phi_{0,0}(z,t) = \int_0^{2\pi} \mathrm{d}\phi\ \int_{-1}^{1} \mathrm{d}\mu\ Y^*_{0,0}(\mu,\phi)\,\Phi(z,\mu,t) = \int_0^{2\pi} \mathrm{d}\phi\ \int_{-1}^{1} \mathrm{d}\mu\ \Phi(z,\mu,t), \tag{5.150}$$

we find that, to calculate these moments, only the coefficients $\Phi_{0,0}$ are needed:

$$\langle z^n(t)\rangle = \frac{\int_{-\infty}^{+\infty} \mathrm{d}z\, z^n\, \Phi_{0,0}(z,t)}{\int_{-\infty}^{+\infty} \mathrm{d}z\, \Phi_{0,0}(z,t)} \equiv \frac{H_0^{(n)}(t)}{H_0^{(0)}(t)}, \tag{5.151}$$

where we introduced the notation

$$H_l^{(n)}(t) = \int_{-\infty}^{+\infty} \mathrm{d}z\, z^n\, \Phi_{l,0}(z,t)\,.$$

We calculate $\langle z^n(t)\rangle$ for $n = 1, 2$ later in this section.

5.4.2 Transverse Displacement

As in the preceding section, we derive the equation for the transverse distribution by, again, integrating Eq. (5.137), but this time over y and z. The integration technique is essentially the same and need not be repeated. The result is the following equation:

$$\frac{\partial}{\partial t}\Phi_{l,m}(x,t) + \sum_{l',m'}\left[\frac{\partial}{\partial x}\Phi_{l',m'}(x,t)\right](Q_{l,m,l',m'})_x + \tilde{\sigma}_l\Phi_{l,m}(x,t) = 0. \tag{5.152}$$

The main difference between this equation and Eq. (5.138) for the longitudinal distribution is that the matrix $(Q\ldots)_z$ is replaced with $(Q\ldots)_x$, where

$$\begin{aligned}(Q_{l,m,l',m'})_x = \int \mathrm{d}\vec{\Omega}\; Y^*_{l,m}\left(\vec{\Omega}\right)\Omega_x Y_{l',m'}\left(\vec{\Omega}\right) = \sqrt{\frac{2l+1}{4\pi}\frac{(l-m)!}{(l+m)!}}\\ \times\sqrt{\frac{2l'+1}{4\pi}\frac{(l'-m')!}{(l'+m')!}}\int_0^{2\pi}\mathrm{d}\phi\;\mathrm{e}^{i(m'-m)\phi}\cos\phi\\ \int_{-1}^{1}\mathrm{d}\mu\; P_l^m(\mu)\sqrt{1-\mu^2}P_{l'}^{m'}(\mu)\end{aligned} \tag{5.153}$$

Here we used $\Omega_x = \sin\theta\cos\phi$. To find this matrix, we first calculate the integral over ϕ:

$$\int_0^{2\pi}\mathrm{d}\phi\;\mathrm{e}^{i(m'-m)\phi}\cos\phi = \pi, \;\text{ if }\; m' = m \pm 1. \tag{5.154}$$

The integral is zero for all other values of m'. Then, we only need to consider $m' = m + 1$ and $m' = m - 1$. To calculate the integral over μ for $m' = m + 1$, we use a recurrence relation for the associated Legendre polynomials (Arfken et al. 2013):

$$\begin{aligned}&\sqrt{1-\mu^2}P_{l'}^{m+1}(\mu)\\&= \frac{1}{2l'+1}\left[\left(l'-m\right)\left(l'-m+1\right)P_{l'+1}^{m} - \left(l'+m\right)\left(l'+m+1\right)P_{l'-1}^{m}\right],\end{aligned} \tag{5.155}$$

and the orthogonality condition:

$$\int_{-1}^{1} \mathrm{d}\mu\, P_l^m(\mu)\, P_k^m(\mu) = \frac{2\,(l+m)!}{(2l+1)\,(l-m)!}\delta_{l,k}. \tag{5.156}$$

We insert Eq. (5.155) into Eq. (5.153), apply the orthogonality condition, and find the matrix elements for $m' = m + 1$:

$$\begin{aligned}\left(Q_{l,m,l',m+1}\right)_x = &\frac{1}{2}\sqrt{\frac{(l-m-1)\,(l-m)}{4l^2-1}}\delta_{l,l'+1}\\&-\frac{1}{2}\sqrt{\frac{(l+m+1)\,(l+m+2)}{(2l+1)\,(2l+3)}}\delta_{l,l'-1}.\end{aligned} \tag{5.157}$$

For $m' = m - 1$ we perform the same procedure, albeit with a different recurrence relation:

$$\sqrt{1-\mu^2}P_{l'}^{m-1}(\mu) = \frac{1}{2l'+1}\left[P_{l'-1}^{m}(\mu) - P_{l'+1}^{m}(\mu)\right]. \tag{5.158}$$

The result is the matrix elements for $m' = m - 1$:

$$\begin{aligned}\left(Q_{l,m,l',m-1}\right)_x = &\frac{1}{2}\sqrt{\frac{(l-m+2)\,(l-m+1)}{(2l+1)\,(2l+3)}}\delta_{l,l'-1}\\&-\frac{1}{2}\sqrt{\frac{(l+m-1)\,(l+m)}{4l^2-1}}\delta_{l,l'+1}.\end{aligned} \tag{5.159}$$

The boundary condition for the fluence is

$$\lim_{t\to 0+}\Phi(x,\mu,t) = \frac{1}{2\pi}\delta(x)\,\delta(\mu-1). \tag{5.160}$$

The boundary condition for Eq. (5.152) follows from the above condition for fluence and Eq. (5.129), which defines the coefficients $\Phi_{l,m}$:

$$\lim_{t\to 0+} \Phi_{l,m}(x,t) = \delta(x)\,\delta_{m,0}\sqrt{\frac{2l+1}{4\pi}}. \tag{5.161}$$

Finally, the moments of the transverse distribution are calculated by this formula:

$$\langle x^n(t)\rangle = \frac{\int_{-\infty}^{+\infty} \mathrm{d}x\, x^n\, \Phi_{0,0}(x,t)}{\int_{-\infty}^{+\infty} \mathrm{d}x\, \Phi_{0,0}(x,t)} \equiv \frac{h_{0,0}^{(n)}(t)}{h_{0,0}^{(0)}(t)}, \tag{5.162}$$

where we defined

$$h_{l,m}^{(n)}(t) = \int_{-\infty}^{+\infty} \mathrm{d}x\, x^n\, \Phi_{l,m}(x,t).$$

Owing to the azimuthal symmetry of the problem, we also have

$$\langle y^n(t)\rangle = \langle x^n(t)\rangle, \tag{5.163}$$

and for the mean-square radial displacement

$$\langle \rho^2(t)\rangle = \langle x^2(t) + y^2(t)\rangle = 2\langle x^2(t)\rangle. \tag{5.164}$$

5.4.3 *Moments of the Longitudinal Distribution*

In this section we calculate the first two moments of the longitudinal spatial distribution using Eq. (5.151). To find the denominator in Eq. (5.151), $H_0^{(0)}$, we set $l = 0$ in Eq. (5.145), and integrate the equation over z, from $-\infty$ to $+\infty$. The integration eliminates the term with the derivative $\partial/\partial z$, because the fluence at infinity is zero. Also, for $l = 0$, we have $\tilde{\sigma}_l = 0$ [see Eq. (5.121)]. This leads to

$$\frac{\partial}{\partial t}\int_{-\infty}^{+\infty} \mathrm{d}z\, \Phi_{0,0}(z,t) = 0. \tag{5.165}$$

Then, by applying the boundary condition, Eq. (5.148), we find:

$$H_0^{(0)}(t) \equiv \int_{-\infty}^{+\infty} \mathrm{d}z\, \Phi_{0,0}(z,t) = \int_{-\infty}^{+\infty} \mathrm{d}z\, \Phi_{0,0}(z,0) = \frac{1}{\sqrt{4\pi}}. \tag{5.166}$$

To find the numerator, $H_0^{(n)}$, for $n = 1$ and $n = 2$, we set $l = 0$ in Eq. (5.145), then multiply the equation by z and z^2, respectively, and, again, integrate over z. For $n = 1$ we have:

$$\frac{\partial}{\partial t}\int_{-\infty}^{+\infty} \mathrm{d}z\, z\, \Phi_{0,0}(z,t) + \int_{-\infty}^{+\infty} \mathrm{d}z\, z\, \frac{\partial}{\partial z}\Phi_{1,0}(z,t)\,\alpha_1 = 0. \tag{5.167}$$

Next, we integrate the second term by parts:

$$\frac{\partial}{\partial t}H_0^{(1)}(t) = \alpha_1 H_1^{(0)}(t). \tag{5.168}$$

This introduces a new unknown, $H_1^{(0)}$. To find it, we set $l = 1$ in Eq. (5.145), and integrate the equation over z. The term with the derivative $\partial/\partial z$ vanishes, resulting in a simple differential equation:

$$\frac{\partial}{\partial t}H_1^{(0)}(t) + \tilde{\sigma}_1 H_1^{(0)}(t) = 0. \tag{5.169}$$

After applying the boundary condition Eq. (5.148), we obtain the solution

$$H_1^{(0)}(t) = \sqrt{\frac{3}{4\pi}}\exp(-\tilde{\sigma}_1 t). \tag{5.170}$$

Inserting this result and $\alpha_1 = 1/\sqrt{3}$ [see Eq. (5.142)] into Eq. (5.168):

$$\frac{\partial}{\partial t}H_0^{(1)}(t) = \frac{1}{\sqrt{4\pi}}\exp(-\tilde{\sigma}_1 t). \tag{5.171}$$

We then integrate, apply the boundary condition $H_0^{(1)}(0) = 0$ that follows from Eq. (5.148), and obtain

$$H_0^{(1)}(t) = \frac{1}{\sqrt{4\pi}\,\tilde{\sigma}_1}\left[1 - \exp(-\tilde{\sigma}_1 t)\right]. \tag{5.172}$$

Finally, we insert $H_0^{(0)}(t)$ and $H_0^{(1)}(t)$ into Eq. (5.151), and find the average longitudinal displacement:

$$\langle z(t)\rangle = \frac{1}{\tilde{\sigma}_1}\left[1 - \exp(-\tilde{\sigma}_1 t)\right]. \tag{5.173}$$

To find $H_0^{(2)}(t)$, required to calculate $\langle z^2\rangle$, we follow a similar procedure. We write Eq. (5.145) for $l = 0$, multiply the equation by z^2, and integrate it over z:

$$\frac{\partial}{\partial t}H_0^{(2)}(t) + \alpha_1\int_{-\infty}^{+\infty} \mathrm{d}z\, z^2\,\frac{\partial}{\partial z}\Phi_{1,0}(z,t) = 0. \tag{5.174}$$

We then integrate the second term by parts:

$$\frac{\partial}{\partial t} H_0^{(2)}(t) = 2\alpha_1 H_1^{(1)}(t) \tag{5.175}$$

This introduces an unknown, $H_1^{(1)}$. To find it we write Eq. (5.145) for $l = 1$, multiply the equation by z, and integrate it over z:

$$\frac{\partial}{\partial t} H_1^{(1)}(t) + \int_{-\infty}^{+\infty} \mathrm{d}z\, z \frac{\partial}{\partial z} \left[\Phi_{0,0}(z,t)\,\alpha_1 + \Phi_{2,0}(z,t)\,\alpha_2\right] + \tilde{\sigma}_1 H_1^{(1)}(t) = 0, \tag{5.176}$$

We then integrate the second term by parts, which yields:

$$\frac{\partial}{\partial t} H_1^{(1)}(t) + \tilde{\sigma}_1 H_1^{(1)}(t) = H_0^{(0)}(t)\,\alpha_1 + H_2^{(0)}(t)\,\alpha_2. \tag{5.177}$$

This introduces yet another unknown, $H_2^{(0)}$. Its derivation is very similar to that of $H_1^{(0)}$, so we simply state the result:

$$H_2^{(0)}(t) = \sqrt{\frac{5}{4\pi}} \exp(-\tilde{\sigma}_2 t). \tag{5.178}$$

We insert this result and $H_0^{(0)}$, given by Eq. (5.166), and $\alpha_1 = 1/\sqrt{3}$, $\alpha_2 = 2/\sqrt{15}$, into the right-hand side of Eq. (5.177):

$$\frac{\partial}{\partial t} H_1^{(1)}(t) + \tilde{\sigma}_1 H_1^{(1)}(t) = \frac{1}{\sqrt{12\pi}} + \frac{1}{\sqrt{3\pi}} \exp(-\tilde{\sigma}_2 t). \tag{5.179}$$

This equation is solved straightforwardly using the method of undetermined coefficients. The solution is sought in the following form:

$$H_1^{(1)}(t) = C(t) \exp(-\tilde{\sigma}_1 t), \tag{5.180}$$

where $C(t)$ is an unknown function that satisfies the boundary condition

$$C(0) = H_1^{(1)}(0) = 0. \tag{5.181}$$

The solution is:

$$H_1^{(1)}(t) = \frac{1 - \mathrm{e}^{-\tilde{\sigma}_1 t}}{\sqrt{12\pi}\,\tilde{\sigma}_1} - \frac{\mathrm{e}^{-\tilde{\sigma}_2 t} - \mathrm{e}^{-\tilde{\sigma}_1 t}}{\sqrt{3\pi}\,(\tilde{\sigma}_2 - \tilde{\sigma}_1)}. \tag{5.182}$$

We insert this result in the right-hand side of Eq. (5.175) for $H_0^{(2)}(t)$, and integrate the equation with the boundary condition $H_0^{(2)}(0) = 0$. This gives us a solution $H_0^{(2)}(t)$.

We then insert the solution in Eq. (5.151) and thus find the second moment of the longitudinal distribution:

$$\langle z^2(t)\rangle = \frac{2}{3}\frac{t}{\tilde{\sigma}_1}\left(1-\frac{1-\mathrm{e}^{-\tilde{\sigma}_1 t}}{\tilde{\sigma}_1 t}\right)+\frac{4}{3}\frac{1}{\tilde{\sigma}_1-\tilde{\sigma}_2}\left(\frac{1-\mathrm{e}^{-\tilde{\sigma}_2 t}}{\tilde{\sigma}_2}-\frac{1-\mathrm{e}^{-\tilde{\sigma}_1 t}}{\tilde{\sigma}_1}\right) \tag{5.183}$$

It can be verified that both the first and the second moments of the longitudinal distribution, Eqs. (5.173) and (5.183) satisfy the obvious condition

$$\lim_{t\to 0}\langle z^n(t)\rangle = 0;\ \ n=1,2. \tag{5.184}$$

The above calculations neglected energy losses, but these can be accounted for by using the continuous slowing down approximation. In this case, the particle energy is a function of the distance t it has traveled, and so are the coefficients $\tilde{\sigma}_l$, (for $l>0$). In this case, the differential equations that we discussed in this section can still be solved, because the main equation of the formalism, Eq. (5.145), is linear first-order. For low-order moments ($n=1,2$) the solution remains relatively straightforward.

5.4.4 *Moments of the Transverse Distribution*

Owing to the azimuthal symmetry, the average displacement of the particle trajectory in the direction of x-axis, $\langle x(t)\rangle$, is zero assuming that the particle started in the z direction. To find $\langle x^2(t)\rangle$, according to Eq. (5.162), we need $h^{(0)}_{0,0}$ and $h^{(2)}_{0,0}$. We start by finding the former, but derive a more general result, $h^{(0)}_{l,m}$. To do so, we integrate Eq. (5.152) over x from $-\infty$ to $+\infty$. The term with the derivative $\partial/\partial x$ vanishes, because the fluence at infinity is zero. The resulting equation is:

$$\frac{\partial}{\partial t}h^{(0)}_{l,m}(t)+\tilde{\sigma}_l h^{(0)}_{l,m}(t)=0. \tag{5.185}$$

The boundary condition for this equation follows from Eq. (5.161):

$$h^{(0)}_{l,m}(0)=\delta_{m,0}\sqrt{\frac{2l+1}{4\pi}}. \tag{5.186}$$

The solution is then:

$$h^{(0)}_{l,m}(t)=\delta_{m,0}\sqrt{\frac{2l+1}{4\pi}}\exp(-\tilde{\sigma}_l t). \tag{5.187}$$

To find $h^{(2)}_{0,0}$, we write Eq. (5.152) for $l = m = 0$, multiply the equation by x^2 and integrate over x:

$$\frac{\partial}{\partial t} h^{(2)}_{0,0}(t) + \sum_{l',m'} \int_{-\infty}^{+\infty} \mathrm{d}x\, x^2 \left[\frac{\partial}{\partial x} \Phi_{l',m'}(x,t)\right] \left(Q_{0,0,l',m'}\right)_x = 0. \tag{5.188}$$

We then integrate the second term by parts. We also note that, according to Eqs. (5.157) and (5.159), the matrix $(Q_{0,0,l',m'})_x$ has only two nonzero elements, at $l' = 1, m' = \pm 1$. This simplifies Eq. (5.188)

$$\frac{\partial}{\partial t} h^{(2)}_{0,0}(t) = 2 \sum_{m'=-1,1} h^{(1)}_{1,m'}(t) \left(Q_{0,0,1,m'}\right)_x, \tag{5.189}$$

but introduces a new unknown, $h^{(1)}_{1,m'}$, $m' = \pm 1$. To find it, we write Eq. (5.152) for $l = 1$, and $m = \pm 1$, then multiply the equation by x, and integrate it over x:

$$\frac{\partial}{\partial t} h^{(1)}_{1,m}(t) + \tilde{\sigma}_1 h^{(1)}_{1,m}(t) = \sum_{l',m'} h^{(0)}_{l',m'}(t) \left(Q_{1,m,l',m'}\right)_x; \;\; m = -1, 1. \tag{5.190}$$

Only two terms in the sum on the right-hand side of this equation are nonzero: $m' = 0$ [see Eq. (5.187)] and $l' = l \pm 1 = 0, 2$ [see Eqs. (5.157) and (5.159)]. The function $h^{(0)}_{l',m'}$ is already known from Eq. (5.187). We find the four coefficients, $(Q_{1,m,0,0})_x$ and $(Q_{1,m,2,0})_x$, $m = \pm 1$ from Eqs. (5.157) and (5.159). This yields two differential equations:

$$\frac{\partial}{\partial t} h^{(1)}_{1,\pm 1}(t) + \tilde{\sigma}_1 h^{(1)}_{1,\pm 1}(t) = \mp \frac{1}{\sqrt{24\pi}} \left[1 - \exp\left(-\tilde{\sigma_2} t\right)\right], \tag{5.191}$$

with the boundary condition that follows from Eq. (5.161):

$$h^{(1)}_{1,\pm 1}(0) = 0. \tag{5.192}$$

We have solved this type of equation previously, see Eq. (5.179). In this case, the solution is

$$h^{(1)}_{1,\pm 1}(t) = \frac{\mp 1}{\sqrt{24\pi}} \left[\frac{1 - \mathrm{e}^{-\tilde{\sigma}_1 t}}{\tilde{\sigma}_1} - \frac{\mathrm{e}^{-\tilde{\sigma}_2 t} - \mathrm{e}^{-\tilde{\sigma}_1 t}}{\tilde{\sigma}_1 - \tilde{\sigma}_2}\right]. \tag{5.193}$$

We can now insert this result into Eq. (5.189) for $h^{(2)}_{0,0}$. Then, after calculating the two coefficients $(Q_{0,0,1,\pm 1})_x$ using Eqs. (5.157) and (5.159), we have

$$\frac{\partial}{\partial t} h^{(2)}_{0,0}(t) = \frac{4}{\sqrt{6}} h^{(1)}_{1,-1}(t), \tag{5.194}$$

where we used $h^{(1)}_{1,1}(t) = -h^{(1)}_{1,-1}(t)$, which is obvious from Eq. (5.193). Next, we insert the expression for $h^{(1)}_{1,-1}(t)$ given in Eq. (5.193) in the right-hand side of Eq. (5.194), and integrate the equation with the boundary condition

$$h^{(2)}_{0,0}(0) = 0. \tag{5.195}$$

The integration gives us the solution $h^{(2)}_{0,0}(t)$. Earlier we also found $h^{(0)}_{0,0}$, Eq. (5.187), which is independent of t because $\tilde{\sigma}_0 = 0$. We insert these two quantities into Eq. (5.162), and thus find the second moment of the transverse distribution:

$$\langle x^2(t)\rangle = \frac{2}{3}\frac{t}{\tilde{\sigma}_1}\left(1 - \frac{1-\mathrm{e}^{-\tilde{\sigma}_1 t}}{\tilde{\sigma}_1 t}\right) + \frac{2}{3}\frac{1}{\tilde{\sigma}_1 - \tilde{\sigma}_2}\left(\frac{1-\mathrm{e}^{-\tilde{\sigma}_1 t}}{\tilde{\sigma}_1} - \frac{1-\mathrm{e}^{-\tilde{\sigma}_2 t}}{\tilde{\sigma}_2}\right). \tag{5.196}$$

5.4.5 *Estimation of Spatial Errors Due to the Finite Step Size*

In condensed history algorithms the step size is a key optimization parameter. In this section we show how to apply the formalism presented above to estimate errors in the spatial distribution of particles as a function of the steps size. We discuss two examples: a very simple basic algorithm, and a much improved algorithm proposed by Larsen (1992).

The Basic Algorithm

In the simplest algorithm, a particle initially departing from the point $\vec{r}_n$ in the direction $\vec{\Omega}_n$, travels a distance t (step size) to arrive at the point:

$$\vec{r}_{n+1} = \vec{r}_n + t\,\vec{\Omega}_n. \tag{5.197}$$

Assuming that $\vec{\Omega}_n$ is parallel to the z-axis, and $x_n = y_n = z_n = 0$, this algorithm yields

$$\langle z\rangle_B = t; \tag{5.198}$$

$$\langle z^2\rangle_B = t^2; \tag{5.199}$$

$$\langle x^2\rangle_B = 0, \tag{5.200}$$

where the subscript B refers to the basic algorithm. We now compare these values with the corresponding exact results $\langle z\rangle$, $\langle z^2\rangle$, and $\langle x^2\rangle$, given by Eqs. (5.173), (5.183), and (5.196). In doing so, we assume that the step size t is

small. More specifically, we assume that $\tilde{\sigma}_l t \ll 1$ for $l = 1, 2$, use Taylor series expansions for the exponents and retain only the first nonvanishing term in the final result. The errors are then:

$$\frac{\langle z\rangle_B - \langle z\rangle}{t} \approx \frac{1}{2}\tilde{\sigma}_1 t; \tag{5.201}$$

$$\frac{\langle z^2\rangle_B - \langle z^2\rangle}{t^2} \approx \frac{1}{3}\tilde{\sigma}_1 t + \frac{2}{9}\tilde{\sigma}_2 t. \tag{5.202}$$

$$\frac{\langle x^2\rangle_B - \langle x^2\rangle}{t^2} \approx -\frac{1}{9}\tilde{\sigma}_2 t. \tag{5.203}$$

These results are qualitatively predictable. By assuming that the particle travels the distance t along a straight line, the basic algorithm overestimates the longitudinal displacement and underestimates the transverse displacement. The errors decrease with decreasing step size.

Knowing the magnitude of the errors, we can modify the algorithm so as to reduce them. An obvious approach is to replace the straight line by a more realistic trajectory. One of the simplest trajectory modifications is to divide t into two straight segments. One such algorithm introduced by Larsen (1992) is presented below as an example.

Larsen's Algorithm

In this algorithm, a particle initially at point $\vec{r}_n$ travels a distance t to arrive at the point $\vec{r}_{n+1}$. However, halfway between the two points, after traveling the distance $t/2$, the particle direction and energy change by scattering. Let the particle parameters at this intermediary point be denoted by the subscript $n + 1/2$. The algorithm is as follows. For completeness, we here include the particle energy.

Algorithm

1. Using the continuous slowing down approximation (CSDA) calculate the particle energy $E_{n+1/2}$ after it has covered distance $t/2$ in the first half of the step.
2. Using the differential scattering cross section for energy $E_{n+1/2}$, sample a new direction $\vec{\Omega}_{n+1/2}$. The scattering angle is sampled using a multiple scattering model, for a particle that has traveled a distance t.
3. Find the particle position at the end of the step:

$$\vec{r}_{n+1} = \vec{r}_n + \frac{t}{2}\vec{\Omega}_n + \frac{t}{2}\vec{\Omega}_{n+1/2}. \tag{5.204}$$

This means that the particle travels the first half of the step in the direction $\vec{\Omega}_n$, and the second half, in the direction $\vec{\Omega}_{n+1/2}$.

4. Using the CSDA, calculate the particle energy E_{n+1} after it has covered the remaining distance $t/2$ of the step.
5. The direction of the particle trajectory at the end of the step is the same as the direction sampled at the midpoint: $\vec{\Omega}_{n+1} = \vec{\Omega}_{n+1/2}$.

Let us now estimate the errors for Larsen's algorithm. We follow the same procedure used for the basic algorithm, neglecting energy losses. The average longitudinal displacement in this model is

$$\frac{\langle z\rangle_L}{t} = \frac{1}{2} + \frac{1}{2}\langle \mu\,(t)\rangle, \tag{5.205}$$

where the subscript L denotes Larsen's algorithm. It is important to note that the scattering angle θ is sampled from the distribution for a particle that has traveled the distance t, and not $t/2$. The average cosine, $\langle \mu(t)\rangle$ is given by Eq. (5.125). Assuming that the step size is small, and using Taylor series expansions for the exponents, we find the error in the average longitudinal displacement for this algorithm:

$$\frac{\langle z\rangle_L - \langle z\rangle}{t} \approx \frac{1}{12}\tilde{\sigma}_1^2 t^2. \tag{5.206}$$

This error is much smaller than that for the basic algorithm, provided that the step size is chosen so that $\tilde{\sigma}_1 t \ll 1$. The error in the second moment $\langle z^2\rangle$ is also lower:

$$\langle z^2\rangle_L = \left\langle \left(\frac{t}{2} + \frac{t}{2}\mu\right)^2 \right\rangle = \frac{t^2}{4}\left[1 + 2\langle \mu\,(t)\rangle + \langle \mu^2\,(t)\rangle\right], \tag{5.207}$$

where we can use Eq. (5.126) for $\langle \mu^2\,(t)\rangle$. The error estimate is then

$$\frac{\langle z^2\rangle_L - \langle z^2\rangle}{t^2} \approx -\frac{1}{6}\tilde{\sigma}_1 t + \frac{1}{18}\tilde{\sigma}_2 t, \tag{5.208}$$

which is smaller than the error of the basic algorithm for any values of the two small parameters $\tilde{\sigma}_1 t$ and $\tilde{\sigma}_2 t$.

Given that the longitudinal displacement estimate has improved, we expect that the transverse displacement estimate is also improved. In the Larsen's model, we have

$$\langle x^2\rangle_L = \left\langle \left(\frac{t}{2}\sin\theta\cos\phi\right)^2 \right\rangle = \frac{t^2}{8}\left[1 - \langle \mu^2\,(t)\rangle\right]. \tag{5.209}$$

Here we used the fact that θ and ϕ are independent random variables. Also, given that the latter is uniformly distributed, the average of $\cos^2\phi$ is 1/2. Following the same procedure as for the longitudinal moments, we find that the error in the transverse distribution is approximately

$$\frac{\langle x^2\rangle_L - \langle x^2\rangle}{t^2} \approx -\frac{1}{36}\tilde{\sigma}_2 t. \tag{5.210}$$

This is four times smaller than the error of the basic algorithm.

The entire above analysis assumed a homogeneous material. Heterogeneities add a further level of complexity to condensed history algorithms. In a voxelized heterogeneous geometry, uncertainties associated with boundary crossings can be limited by reducing the step size when the particle approaches a boundary. The best results are then achieved by switching to the event-by-event algorithm when the particle is close to a boundary. In that case the exact boundary crossing algorithm introduced in Sect. 4.2.2 for neutral particles can be used.

5.5 Heavy Charged Particles in Condensed Matter: Charge Exchange

If a heavy charged particle has some of its electronic shells filled, for example, a carbon ion C^{4+}, then through interaction with matter it may lose some of the electrons and transition to another ionized state (e.g., $C^{4+} \rightarrow C^{5+}$) or become fully ionized ($C^{4+} \rightarrow C^{6+}$). Alternatively, free electrons in the medium can bind to the ion and thus reduce its charge (e.g., $C^{4+} \rightarrow C^{3+}$). These two opposite processes, electron loss and capture, are called charge exchange. This process is particularly important at low energies, near the track end.

Definitions

- The charge state i of an ion is the difference of its atomic number Z_1 and the number of electrons in bound states of the ion.
- The charge state distribution, Φ_i, is the fraction of ions in charge state i; Φ_i is the fluence normalized so that $\sum_i \Phi_i = 1$.
- The fractional charge q, mean charge $\bar{i}$, and the width d of the charge state distribution are defined by the following equations:

$$q = \frac{i}{Z_1}. \tag{5.211}$$

$$\bar{i} = \sum_i i\Phi_i. \tag{5.212}$$

$$d = \sqrt{\sum_i \left(i - \bar{i}\right)^2 \Phi_i}. \tag{5.213}$$

Equation for Charge State Distribution

As particles travel in matter, their charge state distribution changes as a function of penetration depth, x. Evolution of charge state distribution is described by the following equation (Bridwell et al. 1986)

$$\frac{\mathrm{d}\Phi_i}{\mathrm{d}x} = \sum_{i' \neq i} \left(\sigma_{i'i}\Phi_{i'} - \sigma_{ii'}\Phi_i\right), \tag{5.214}$$

where $\sigma_{i'i}$ is the cross section for transition from charge state i' to charge state i. At a certain depth, an equilibrium charge state distribution is reached that is independent of the initial charge states of the ions at $x = 0$. The equilibrium depth, x_∞, is determined by the charge exchange cross sections. Baron (1979) reported the following simple estimate for x_∞ derived as a fit to experimental data for solids, mostly carbon:

$$x_\infty = 5.9 + 22.4E - 1.13E^2, \tag{5.215}$$

where the depth x_∞ is in units of $\mu\mathrm{g/cm}^2$ and ion energy E is in units of MeV/nucleon. The experimental data was limited to $E \lesssim 10$ MeV/nucleon. In gases, x_∞ tends to be lower than that in solids (Zaikov et al. 1984). From Eq. (5.215) it can be estimated that the equilibrium depth is of the order of micrometers. Hence, most ion beam experiments are performed under charge state equilibrium conditions. An exception is experiments where ions scattered by the surface layer of the target are registered. Further, we consider only equilibrium conditions. In this case, evolution of the charge state distribution is driven only by slowing down of the ions.

Slow Ions

In this context, ions with velocity v_1 that is much lower than the Fermi velocity v_F of the target material are categorized as "slow." For hydrocarbon targets, for example, the Fermi velocity can be approximated by the Bohr velocity $v_0 \approx 2.2 \times 10^8$ cm/s (Sabin and Brandas 2004). This is the velocity of an ion with kinetic energy of 25 keV/nucleon.

At these low velocities the main factor determining the charge state of an ion is screening of the ion charge by electrons of the target (Echenique et al. 1988). Bargmann (1952) has shown that the number of bound states n_{bl} with angular momentum l in a central field $V(r)$ is limited according to the following inequality:

$$(2l + 1)\, n_{bl} < \int_0^\infty \mathrm{d}r\, r|V(r)| \tag{5.216}$$

Using the screened Coulomb potential for $V(r)$, Rogers et al. (1970) found that the number of stable bound states is approximately

$$n^* = 0.5829 + 0.4993\, aZ_1, \tag{5.217}$$

where a is the screening length in atomic units. It can be estimated using the free electron gas model. In this model the screening length is determined by the electron density n_e of the target, and can be expressed in terms of the radius r_s of a sphere that contains one electron on average as (Kaneko 1986)

$$a = \left(\frac{\pi}{12}\right)^{1/3} r_s^{1/2};\ \ r_s = \left(\frac{4}{3}\pi n_e\right)^{-1/3}. \tag{5.218}$$

Simple calculations show that in most solids helium ions can have no more than one bound electron (Kaneko 1986). A proton cannot bind an electron in solids with $r_s < 1.7$ (in atomic units).

Ions of High and Intermediate Energies

Ions in this category have a velocity exceeding the Fermi velocity of the target material, i.e., $v_1 \gtrsim v_F$. At these velocities, Bohr's criterion is applicable. According to the criterion (Datz et al. 1975; Gervasoni and Cruz-Jimenez 1996), ions traveling with velocity v_1 in condensed matter lose all those electrons whose orbital velocities are less than v_1. In a hydrogen-like ion, the orbital velocity of K-shell electrons is $v_0 Z_1$, where v_0 is the Bohr velocity. Then, ions traveling at velocities $v_1 > v_0 Z_1$ are fully stripped of electrons, and the fractional charge is q=1 (Datz et al. 1975).

For ion velocities comparable to the Fermi velocity v_F of the target material, a refined version of Bohr's criterion has been proposed (Kreussler et al. 1981; Brandt and Kitagawa 1982): ions traveling with velocity v_1 in condensed matter lose all those electrons whose orbital velocities are less than the relative velocity v_r of the ions with respect to electrons of the medium. Assuming that all electrons in the medium have the same velocity v_e and an isotropic distribution of their directions, the average relative velocity is (Kreussler et al. 1981):

$$\bar{v}_r = \frac{v_e}{6}\left[\left(\frac{v_1}{v_e}+1\right)^3 - \left|\frac{v_1}{v_e}-1\right|^3\right]\frac{v_e}{v_1}. \tag{5.219}$$

A more realistic model for the distribution of electron velocity is the free electron gas model. In this model, electron velocity $\vec{v}_e$ (velocity vector) is distributed

uniformly in a sphere of a radius equal to the Fermi velocity of the material, v_F. In this model, averaging over the electron velocity distribution yields

$$\bar{v}_r = \begin{cases} v_1 \left[1 + \frac{1}{5}\left(v_F/v_1\right)^2\right], & \text{if } v_1 \geq v_F; \\ \frac{3}{4} v_F \left[1 + \frac{2}{3}\left(v_1/v_F\right)^2 - \frac{1}{15}\left(v_1/v_F\right)^4\right], & \text{if } v_1 < v_F. \end{cases} \tag{5.220}$$

Transition Effects

A charge state distribution of ions when the ions travel in condensed matter does not necessarily remain the same after they exit into air or vacuum, where they are usually detected. Two reasons for changes in charge state distributions at the transition point have been suggested in the literature:

- As a result of the interaction with the Coulomb field of the ion, a free electron in the medium may acquire the same velocity as the ion and escort it for some distance. This process is called electron capture into continuum. Once the ion exits into air, the captured electron may transition into a bound state because screening of the ion charge by electrons of the medium suddenly drops (Datz et al. 1975).
- An ion traveling in condensed matter constantly interacts with the medium, which brings its electron shell to an excited state, for significant periods of time. When the ion exits condensed matter, de-excitation processes begin, that may result in the emission of an Auger electron (Kaneko 1986; Shima et al. 1988).

5.6 Transport of Charged Particles in Magnetic Fields

Particle Trajectory in Vacuum

In the presence of a magnetic field $\vec{B}$ a particle with charge q, traveling with velocity $\vec{v}$ is subject to the Lorentz force (Jackson 1999):

$$\vec{F}_L = q\vec{v} \times \vec{B}. \tag{5.221}$$

Accordingly, the equation of motion is

$$\frac{\mathrm{d}\vec{p}}{\mathrm{d}t} = q\vec{v} \times \vec{B} \tag{5.222}$$

To replace the particle momentum $\vec{p}$ with velocity v, we use the relativistic formula

$$\vec{p} = \frac{E}{c^2}\vec{v}, \tag{5.223}$$

where E is the total energy of the particle, including the rest energy. In a magnetic field, the particle energy is constant, if no other forces are applied. With the notation

$$\vec{\omega}_B \equiv \frac{qc^2}{E}\vec{B} \tag{5.224}$$

the equation of motion takes a simpler form

$$\frac{\mathrm{d}\vec{v}}{\mathrm{d}t} = \vec{v} \times \vec{\omega}_B. \tag{5.225}$$

The quantity $|\vec{\omega}_B|$ is called the gyration or precession frequency. Its units are rad/s. The initial condition for this differential equation is

$$\vec{v}\,(0) = \vec{v}_0. \tag{5.226}$$

The magnetic field generally is a function of time t and position $\vec{r}$. In this case, the equation of motion can be solved numerically using one of the standard techniques, such as the Runge–Kutta method. Here, we present an analytical solution for a spatially uniform and time-independent magnetic field. We will not consider electric fields. The electric field is set to zero.

We choose the coordinate system so that the z-axis is parallel to the magnetic field $\vec{B}$. In that case $\omega_{Bx} = \omega_{By} = 0$, $\omega_{Bz} = \omega_B$, and the equation of motion in componentwise form is

$$\begin{aligned}
\frac{\mathrm{d}v_x}{\mathrm{d}t} &= v_y\omega_B. \\
\frac{\mathrm{d}v_y}{\mathrm{d}t} &= -v_x\omega_B. \\
\frac{\mathrm{d}v_z}{\mathrm{d}t} &= 0.
\end{aligned} \tag{5.227}$$

Because neither the particle energy E nor magnetic field $\vec{B}$ change with time, the gyration vector $\vec{\omega}_B$ is constant. Then, the solution of the above system of equations satisfying the boundary condition Eq. (5.226) is

$$\begin{aligned}
v_x &= v_{0x}\cos(\omega_B t) + v_{0y}\sin(\omega_B t). \\
v_y &= -v_{0x}\sin(\omega_B t) + v_{0y}\cos(\omega_B t). \\
v_z &= v_{0z}.
\end{aligned} \tag{5.228}$$

The first two equations are the equation of a circle in the (v_x, v_y) plane. The length $|\vec{v}|$ of the velocity vector is conserved.

To find the particle trajectory, $\vec{r}(t)$, we integrate Eq. (5.228) with the initial condition

$$\vec{r}\,(0) = \vec{r}_0. \tag{5.229}$$

The solution is the equation of a helix in parametric form

$$\begin{aligned} x &= x_0 + \frac{v_{0y}}{\omega_B} + \frac{v_{0x}}{\omega_B}\sin(\omega_B t) - \frac{v_{0y}}{\omega_B}\cos(\omega_B t). \\ y &= y_0 - \frac{v_{0x}}{\omega_B} + \frac{v_{0x}}{\omega_B}\cos(\omega_B t) + \frac{v_{0y}}{\omega_B}\sin(\omega_B t). \\ z &= z_0 + v_{0z}t. \end{aligned} \tag{5.230}$$

The radius of the helix is

$$R_B = \frac{1}{\omega_B}\sqrt{v_{0x}^2 + v_{0y}^2}. \tag{5.231}$$

For example, for electrons in a 1.5 T magnetic field, the helix radius is 3.4 mm and 10 mm at electron kinetic energies of 1 MeV and 4 MeV, respectively (Raaijmakers et al. 2007).

Event-by-Event Algorithms

In event-by-event algorithms a charged particle does not interact with matter as it travels a distance l between two consecutive collisions called the free path. Then, in a magnetic field, the distance l is traveled along a helical trajectory defined by Eq. (5.230). At the end of the free path, an interaction is simulated and, unless the particle is absorbed, the next free path is sampled. The particle trajectory continues along a different helix, because the particle momentum changes in the collision. The boundary crossing algorithm does not change in a magnetic field, except that the distance to the boundary is calculated using the equation of the helical trajectory instead of the equation of a straight line. Alternatively, the step size can be limited so that the particle travels a distance l in small steps Δl, with each step being a straight line segment approximating the helix. This approach introduces an error and should only be used when an analytical solution of the equation of motion cannot be found.

Condensed History Algorithms

Developing an algorithm should start with choosing an appropriate form of the Boltzmann equation. The Boltzmann equation in the presence of external magnetic

fields was discussed in Chap. 3. It was derived recently by St. Aubin et al. (2015, 2016) and Bouchard and Bielajew (2015). These studies were motivated by the recent advances in MRI-guided radiotherapy, because clinical implementation of this new technology requires dose calculations in the presence of magnetic fields. However, a Monte Carlo algorithm for solving this form of the Boltzmann equation has not been developed yet. St. Aubin et al. (2015, 2016) did not use the Monte Carlo method. Instead, they took an entirely different approach. They developed an algorithm and solved the equation with a grid-based solver, which is a deterministic method that we discuss in Chap. 7. Bouchard and Bielajew (2015) discussed general implications of having the Lorentz force term in the equation, but did not offer an algorithm for solving it.

The Monte Carlo algorithm that has been used for transport of charged particles in magnetic fields is a combination of the mechanistic approach for modeling a helical particle trajectory in a magnetic field that we described above, with the statistical approach that uses multiple scattering models. The algorithm was introduced in Jenkins et al. (1988). A good description of the algorithm is also given in Salvat et al. (2011), Keyvanloo et al. (2012), and Yang and Bednarz (2013).

In condensed history algorithms the concept of a free path is not applicable. Instead, as we have discussed in this chapter, the trajectory of a charged particle is generated step-by-step, with each step accounting for multiple interactions. The principal difficulty of using algorithms of this type in magnetic fields is that particle momentum changes continuously and neither particle energy nor the direction of its velocity are known at each point of the trajectory. They are sampled only at the end of the step. The Lorentz force, of course, depends both on particle energy and direction it is traveling.

To account for all scattering that a particle undergoes as it travels the step length, we add to the equation of motion Eq. (5.222) a force $\vec{F}$

$$\frac{\mathrm{d}\vec{p}}{\mathrm{d}t} = q\vec{v} \times \vec{B} + \vec{F}(v). \tag{5.232}$$

The force accounts for energy losses and deflections resulting from particle interactions with matter. It is a macroscopic force, representing the average effect of all interactions of the particle with atoms along its path. The force is not explicitly known. We only know through multiple scattering theories, its net effect on the distribution of particle momentum at the end of the step.

We rewrite Eq. (5.232) using

$$\vec{p} = \frac{m\vec{v}}{\sqrt{1 - v^2/c^2}} \equiv m\gamma(v)\,\vec{v}. \tag{5.233}$$

to eliminate the momentum from the equation:

$$m\frac{\mathrm{d}}{\mathrm{d}t}\left[\gamma(v)\,\vec{v}\,\right] = q\vec{v} \times \vec{B} + \vec{F}(v). \tag{5.234}$$

Then, we choose a sufficiently small step size l so that when integrating Eq. (5.234) we can assume that the right-hand side of the equation and $\gamma(v)$ are constant:

$$m\gamma\left(v_0\right)\frac{\mathrm{d}\vec{v}}{\mathrm{d}t}=q\vec{v}_0\times\vec{B}+\vec{F}\left(v_0\right). \tag{5.235}$$

Here $\vec{v}_0$ is the initial velocity, i.e., the particle velocity at the beginning of the step, at $t=0$. We can easily integrate this equation and find the change of the particle velocity at the end of the step:

$$\Delta\vec{v}\left(l\right)\equiv\vec{v}\left(t\right)-\vec{v}_0=\vec{v}_0\times\vec{\omega}_B\left(v_0\right)t+\frac{\vec{F}\left(v_0\right)t}{m\gamma\left(v_0\right)}. \tag{5.236}$$

Finally, as a matter of convenience, we express the right-hand side in terms of the step length, instead of time t. To do so, we use the obvious formula

$$t=\int_0^l\frac{\mathrm{d}l'}{v\left(l'\right)}\approx\frac{l}{v_0}. \tag{5.237}$$

Furthermore, we prefer to use the directional vector $\vec{\Omega}$ rather than velocity $\vec{v}$. Hence, we write the final result as

$$\Delta\vec{\Omega}\left(l\right)=\frac{\Delta\vec{v}\left(l\right)}{v_0}=\vec{\Omega}_0\times\vec{\omega}_B\left(v_0\right)\frac{l}{v_0}+\frac{\vec{F}\left(v_0\right)l}{m\gamma\left(v_0\right)v_0^2}\equiv\Delta\vec{\Omega}_L\left(l\right)+\Delta\vec{\Omega}_F\left(l\right). \tag{5.238}$$

Still, the force $\vec{F}(v_0)$ remains unknown. The meaning of Eq. (5.238) is that the magnetic field simply adds deflection $\Delta\vec{\Omega}_L\left(l\right)$ to the particle trajectory. This observation justifies the following algorithm. First, a step is made as if the magnetic field was not present. The particle energy and deflection $\Delta\vec{\Omega}_F\left(l\right)$ at the end of the step are sampled using one of the standard multiple scattering models. Then the deflection caused by the magnetic field $\Delta\vec{\Omega}_L\left(l\right)$ is added. The magnetic field does not change the particle energy. Because Eq. (5.238) does not preserve vector length, the directional vector $\vec{\Omega}$ has to be renormalized before starting the next step:

$$\vec{\Omega}\left(l\right)=\frac{\vec{\Omega}_0+\Delta\vec{\Omega}\left(l\right)}{|\vec{\Omega}_0+\Delta\vec{\Omega}\left(l\right)|} \tag{5.239}$$

As we discussed in this chapter, multiple scattering models impose restrictions on the step size l. The algorithm in a magnetic field that we have just presented, introduces additional restrictions. First, the velocity change should be small, i.e.,

$$\frac{|\Delta\vec{v}\left(l\right)|}{v_0}\ll1. \tag{5.240}$$

Alternatively, the above requirement can be expressed in terms of energy loss.

Second, even though we assumed that the magnetic field was constant, the algorithm is also applicable to spatially nonuniform fields. In that case we require that the change of the magnetic field over a step length is small. For any point $\vec{r}$ on the particle path l the following inequality must be satisfied

$$\frac{|\vec{B}(\vec{r}) - \vec{B}_0|}{|\vec{B}_0|} \ll 1, \tag{5.241}$$

where $\vec{B}_0$ is the magnetic field at the beginning of the step.

Third, the deflection caused by the magnetic field must be small, i.e.

$$\left|\Delta\vec{\Omega}_L\right| = \left|\vec{\Omega}_0 \times \vec{\omega}_B(v_0)\frac{l}{v_0}\right| \ll 1. \tag{5.242}$$

This is because we used an approximate solution of the equation of motion to find $\Delta\vec{\Omega}_L$.

These additional constraints are rather restrictive and may slow down calculations significantly, especially in strong magnetic fields. Kirkby et al. (2010), for example, reported the maximum step sizes of 0.2, 0.02, and 0.01 cm for electrons in magnetic fields of 0.2 T, 1.5 T, and 3.0 T, respectively.

References

Abramowitz, M., Stegan I.A. (eds.): Handbook of Mathematical Functions: with Formulas, Graphs, and Mathematical Tables. Dover Publications, New York (1964)

Aghili, A., Moghaddam, B.P.: Certain theorems on two dimensional Laplace transform and non-homogeneous parabolic partial differential equations. Surv. Math. Appl. **6**, 165–174 (2011)

Arfken, G.B.: Mathematical Methods for Physicists, 3rd edn. Academic, New York (1985)

Arfken, G.B., Weber, H.J., Harris, F.E.: Mathematical Methods for Physicists, 7th edn. Elsevier, Amsterdam (2013)

Bargmann, V.: On the number of bound states in a central field of force. Proc. Natl. Acad. Sci. **38**(11), 961–966 (1952)

Baron, E.: Beam-stripper interaction studies for GANIL. IEEE Trans. Nucl. Sci. **26**(2), 2411–2413 (1979)

Bernal, M.A., Bordage, M.C., Brown, J.M.C., Davidkova, M., Delage, E., El Bitar, Z., Enger, S.A., Francis, Z., Guatelli, S., Ivanchenko, V.N., Karamitros, M., Kyriakou, I., Maigne, L., Meylan, S., Murakami, K., Okada, S., Payno, H., Perrot, Y., Petrovic, I., Pham, Q.T., Ristic- Fira, A., Sasaki, T., Stepan, V., Tran, H.N., Villagrasa, C., Incerti, S.: Track structure modeling in liquid water: a review of the Geant4-DNA very low energy extension of the Geant4 Monte Carlo simulation toolkit. Phys. Med. **31**(8), 861–874 (2015)

Bethe, H.A.: Molière's theory of multiple scattering. Phys. Rev. **89**(6), 1256–1266 (1953)

Brandt, W., Kitagawa, M.: Effective stopping-power charges of swift ions in condensed matter. Phys. Rev. B **25**(9), 5631–5637 (1982)

Bridwell, L.B., Cowern, N.E.B., Read, P.M., Sofield, C.J.: Measurement of polarization, Bloch and charge exchange contributions to the stopping power of C and O ions in carbon. Nucl. Instrum. Meth. B **13**(1–3), 123–126 (1986)

Bouchard, H., Bielajew, A.: Lorentz force correction to the Boltzmann radiation transport equation and its implications for Monte Carlo algorithms. Phys. Med. Biol. **60**(13), 4963–4971 (2015)

Datz, S., Appleton, B.R., Moak, C.D. (eds.): Atomic Collisions in Solids. Springer, Berlin (1975)

Echenique, P.M., Flores, F., Ritchie, R.H.: Dynamic screening: capture and loss processes of protons moving in solids. Nucl. Instrum. Meth. B **B33**(1–4), 91–97 (1988)

Gervasoni, J.L., Cruz-Jimenez, S.: Bohr's adiabatic criterion and effective charge of heavy ions. Radiat. Phys. Chem. **48**(4), 433–436 (1996)

Goudsmit, S., Saunderson, J.L.: Multiple scattering of electrons. Phys. Rev. **57**(1), 24–29 (1940a)

Goudsmit, S., Saunderson, J.L.: Multiple scattering of electrons. II. Phys. Rev. **58**, 36–42 (1940b)

Gradshteyn, I.S., Ryzhik, I.M.: Table of Integrals, Series and Products, Corrected and Enlarged Edition. Academic, San Diego (1980)

Incerti, S., Baldacchino, G., Bernal, M., Capra, R., Champion, C., Francis, Z., Guatelli, S., Guye, P., Mantero, A., Mascialino, B., Moretto, P., Nieminen, P., Villagrasa, C., Zacharatou, C.: The Geant4-DNA project. Int. J. Model. Simul. Sci. Comput. **1**, 157–178 (2010)

Jackson, J.D.: Classical Electrodynamics. 3rd edn. Wiley, New York (1999)

Jenkins, T.M., Nelson, W.R., Rindi, A., Nahum, A.E., Rogers, D.W.O., Bielajew, A.F. (eds.): Monte Carlo Transport of Electrons and Photons. Plenum Press, New York (1988)

Kaneko, T: Energy loss of protons and helium ions passing through matter. Phys. Rev. A **33**(3), 1602–1611 (1986)

Keyvanloo, A., Burke, B., Warkentin, B., Tadic, T., Rathee, S., Kirkby, C., Santos, D.M., Fallone, B.G.: Skin dose in longitudinal and transverse linac-MRIs using Monte Carlo and realistic 3D MRI field models. Med. Phys. **39**(10), 6509–6521 (2012)

Kirkby, C., Murray, B., Rathee, S., Fallone, B.G.: Lung dosimetry in a linac-MRI radiotherapy unit with a longitudinal magnetic field. Med. Phys. **37**(9), 4722–4732 (2010)

Kreussler, S., Varelas, C., Brandt, W.: Target dependence of effective projectile charge in stopping powers. Phys. Rev. B **23**(1), 82–84 (1981)

Landau, L.: On the energy loss of fast particles by ionisation. J. Phys. **8**(1–6):201–205 (1944)

Larsen, E.W.: A theoretical derivation of the condensed history algorithm. Ann. Nucl. Energy **19**(10–12), 701–714 (1992)

LePage, W.R.: Complex Variables and the Laplace Transform for Engineers. Dover Publications, New York (2010)

Lewis, H.W.: Multiple scattering in an infinite medium. Phys. Rev. **78**(5), 526–529 (1950)

Macdonald, H.M.: The Addition theorem for the Bessel functions. Proc. Lond. Math. Soc. **s1–32**(1):152–157 (1900)

Molière, G.: Theorie der Streuung schneller geladener Teilchen I. Einzelstreuung am abgeschirmten Coulomb-Feld. Z. Naturforsch A **2**(3), 133–145 (1947)

Molière, G.: Theorie der Streuung schneller geladener Teilchen II. Mehrfach- und Vielfachstreuung. Z. Naturforsch A **3**(2), 78–97 (1948)

Nikjoo, H., Uehara S., Emfietzoglou D., Brahme A.: Heavy charged particles in radiation biology and biophysics. New J. Phys. **10**, 075006 (2008)

Poularikas, A.D. (ed.): Transforms and Applications Handbook. 3rd edn. CRC Press, Boca Raton, FL (2010)

Raaijmakers, A.J.E., Raaymakers, B.W., van der Meer, S., Lagendijk, J.J.W.: Integrating a MRI scanner with a 6 MV radiotherapy accelerator: impact of the surface orientation on the entrance and exit dose due to the transverse magnetic field. Phys. Med. Biol. **52**(4), 929–939 (2007)

Rogers, F.J., Graboske, H.C., Harwood, D.J.: Bound eigenstates of the static screened Coulomb potential. Phys. Rev. A **1**(6), 1577–1586 (1970)

Rutherford, E.: The scattering of α and β particles by matter and the structure of the atom. Philos. Mag. **21**(125), 669–688 (1911)

Sabin, J.R., Brandas, E. (eds.): Advances in Quantum Chemistry. Theory of Interactions of Swift Ions with Matter. Part 2. Elsevier, Amsterdam (2004)

Salvat, F., Fernandez–Varea, J.M., Sempau, J.: PENELOPE-2011. A code system for Monte Carlo simulation of electron and photon transport: Nuclear Energy Agency (2011)

Shima, K., Nakagawa E., Kakita T., Yamanouchi M, Awaya Y., Kambara T., Mizogawa T., Kanai Y.: Projectile atomic number dependence of equilibrium charge states for 1 and 2 MeV ions passing through a carbon foil. Nucl. Instrum. Meth. B **B33**(1–4), 212–215 (1988)

St. Aubin, J., Keyvanloo, A., Vassiliev, O., Fallone, B.G.: A deterministic solution of the first order linear Boltzmann transport equation in the presence of external magnetic fields. Med. Phys. **42**(2), 780–793 (2015)

St. Aubin, J., Keyvanloo, A., Fallone, B.G.: Discontinuous finite element space-angle treatment of the first order linear Boltzmann transport equation with magnetic fields: application to MRI-guided radiotherapy. Med. Phys. **43**(1), 195–204 (2016)

Vavilov, P.V.: Ionization losses by high-energy heavy particles. Sov. Phys. JETP **5**(4), 749–751 (1957)

Wang, H., Vassiliev, O.N.: Microdosimetric characterisation of radiation fields for modelling tissue response in radiotherapy. Int. J. Cancer Ther. Oncol. **2**(1), 020116 (2014)

Yang, Y.M., Bednarz, B.: Consistency evaluation between EGSnrc and Geant4 charged particle transport in an equilibrium magnetic field. Phys. Med. Biol. **58**(4), N47–N58 (2013)

Zaikov, V.P., Kral'kina, E.A., Vorobjev, N.F., Dmitriev, I.S., Nikolaev, V.S., Teplova, Y.A.: Attainment of equilibrium charge distributions in fast ion beams passing through solid films. Nucl. Instr. Meth. B **5**(1), 10–13 (1984)

Chapter 6
Microdosimetry. Elements of Stochastic Transport Theory

6.1 Beyond the Boltzmann Equation

So far, in this book, we have been mostly concerned with calculation of macroscopic characteristics of radiation fields, such as dose distributions on a spatial scale on the order of ~ 1 mm. However, absorbed dose does not always serve as a good predictor of biological effects of radiation, including the outcome of a radiation treatment. It is well known that depending on the type of radiation, the same dose to the same biological system can have very different biological sequelae. The effects can be different even for the same dose and same radiation type, but with two beams having different energy spectra. Understanding and quantifying these differences is becoming increasingly important, owing largely to the rapid expansion of hadron therapy.

The physical mechanisms underlying different biological responses to radiation are related to the properties of dose distributions on a microscopic scale that is comparable to the size of subcellular structures, ranging from the cell nucleus to the DNA helix. The relevant scale is from a few nanometers to a few micrometers. Strictly speaking, the absorbed dose, by definition, is a macroscopic quantity and on this microscopic scale, the energy deposited by radiation should be used instead. The deposited energy is a stochastic quantity. Its microscopic spatial distribution is extremely heterogeneous and is best described in statistical terms. Microdosimetry offers an established formalism that is well suited for that purpose.

The calculation of the statistical characteristics of radiation fields poses a challenge for Monte Carlo simulations. The fundamental difficulty is that the Boltzmann equation is not applicable to fluctuational characteristics of radiation fields. It is an equation for fluence. By solving the Boltzmann equation we can find, for example, the average energy deposited in a volume. If, on the other hand, we are interested in

O.N. Vassiliev, *Monte Carlo Methods for Radiation Transport*, Biological and Medical Physics, Biomedical Engineering, DOI 10.1007/978-3-319-44141-2_6

the variance of the deposited energy, solving the Boltzmann equation will not give us the answer. It simply does not contain this type of information. One approach to circumvent this problem is by using intuitive algorithms. Several such algorithms are presented in this chapter. The Monte Carlo method is an inherently statistical method, and as we generate particle trajectories, we can certainly calculate not only the average deposited energy, but also the variance of that quantity. However, we cannot use powerful variance reduction techniques and other numerical techniques that ultimately are based on the properties of the Boltzmann equation. This is a serious drawback of the intuitive approach, considering further complicating factors associated with this numerical problem. These additional factors will be discussed below.

Clearly, for efficiently solving this category of problems, there is a need to devise methods beyond the Boltzmann equation. In this section we introduce an approach, the fluctuation detector (FD) method, which is based on the principles of stochastic radiation transport theory. The theory dates back to the mid-sixties. It was developed primarily for studying fluctuational aspects of neutron transport (Bell 1965; Saito and Otsuka 1965). A form of the stochastic transport equation will be presented later in this chapter. It is a nonlinear integro-differential equation, somewhat similar to the adjoint transport equation [Sect. 3.3, Eq. (3.48)].

6.2 Event-by-Event Monte Carlo Algorithms

Monte Carlo simulations have been used extensively for the analysis of microscopic patterns of energy deposition, as well as for the modeling of biological effects on the cellular level, such as DNA damage. In this type of calculations best results are achieved with the so-called event-by-event algorithms, also known as track structure algorithms. In event-by-event algorithms, all interactions such as ionizations, excitations, and elastic scattering, are simulated explicitly. This is in contrast to condensed history algorithms in which the step size exceeds the mean free path, and at the end of the step the net result of multiple collisions is simulated. The advantage of event-by-event algorithms is that they provide maximal spatial resolution of microscopic energy deposition patterns, or track structure. They are typically designed so that very low tracking cut-off can be used, on the order of 10 eV for electrons, which is comparable to the ionization threshold. Unavoidably, event-by-event algorithms are slow. Fortunately, for many applications it suffices to model only a relatively short track segment, for example a track segment just exceeding the size of the cell nucleus. Algorithms of this type were reviewed by Nikjoo et al. (2006). We will add to the list an important new code Geant4-DNA (Incerti et al. 2010; Bernal et al. 2015), that was not available at the time of the review.

A major challenge for event-by-event algorithms are large uncertainties that exist in particle interaction parameters, such as the cross sections and the stopping power, at low energies. Mostly for this reason, the software is usually written for one

material, water. Earlier software used interaction data for water vapor, scaled to the density of liquid water. However, it is now well understood that at low energies, the phase effects are not negligible. Interaction parameters for liquid water are significantly different from those derived by density scaling of parameters for water vapor, see for example Turner et al. (1982). Thus research in this area has shifted more towards developing interaction models and simulation algorithms for liquid water.

Another complication associated with low energies is modeling processes that are neglected or not explicitly modeled in radiotherapy dose calculations. Examples of such processes are excitation of molecular vibrational degrees of freedom, charge exchange processes for heavy charged particles (see Sect. 5.5), and emission of Auger electrons following ionization of inner shells. In tissue and in water, the energy of Auger electrons is on the order of a few hundred electron volts, which is higher than the typical tracking cut-off in event-by-event algorithms.

Some of the more advanced software systems are not limited to radiation transport. They may include the simulation of radiolysis, the production of radical species, their diffusion and their interactions with each other and with DNA molecules, and the formation of DNA damage. Some of these systems include sophisticated models of the DNA geometry, and some are interfaced with molecular dynamics software.

6.3 Microdosimetry

As mentioned above, microscopic distributions of deposited energy should be described in statistical terms. In this section we present a well-established framework that is well suited for that purpose. The framework includes a set of physical quantities, methods for their measurements and calculations, and a formalism for translating physical data into estimates of expected biological effects. The framework forms the basis of what we would describe as classical microdosimetry. In a more general sense, microdosimetry is a very broad field that includes a variety of approaches towards studying radiation effects. The foundations of microdosimetry were presented in a series of papers published in the mid-fifties by H. Rossi and W. Rosenzweig. The most important contribution was the development of a device for measuring probability distributions of energy deposited in micrometer-sized volumes (Rossi and Rosenzweig 1955). This device is now known as the Rossi counter or the proportional counter. The foundations of microdosimetry and its applications are discussed in depth in Rossi and Zaider (1996) and ICRU 36 (1983).

6.3.1 *Definitions*

Microdosimetry

Microdosimetry is defined, rather formally, by Rossi and Zaider (1996), as the systematic study and quantification of the *spatial* and *temporal* distributions of absorbed energy in irradiated matter.

Deposited Energy, ϵ

The exact meaning of ϵ may depend on the application, radiobiological model, and the method of measurement. In condensed matter, for example, the situation is complicated by the fact that the energy is delocalized through collective excitations, or quasiparticles, such as plasmons. Therefore, we will leave this term open to interpretation and will refrain from offering a rigorous definition.

The Sensitive Volume

The sensitive volume (SV) is the volume of interest. The term was originally intended to represent the cell nucleus. A typical SV in classical microdosimetry is a sphere of a tissue equivalent material, 1 μm in diameter. More recently, interest has apparently shifted to the nanoscale, with SVs representing DNA molecules.

The Probability Distribution of Deposited Energy, ϵ

The probability distribution of the total energy deposited in the SV by all particles emitted by a source, $\epsilon = \epsilon_1 + \epsilon_2 + \ldots + \epsilon_\nu$, depends on the total delivered dose, D. We will denote this distribution by $f(q|D)$. This distribution has a singularity at $q = 0$, because the probability of zero energy deposited, $\epsilon = 0$ is finite:

$$P(\epsilon = 0|D) = 1 - P(\epsilon > 0|D) = 1 - \int_{0+}^{\infty} f(q|D)\,\mathrm{d}q, \tag{6.1}$$

where $0+$ means that the point $q = 0$ is not included in the integral. For nanoscale SVs at therapeutic doses, $P(\epsilon > 0|D) \ll 1$. For example, for proton beams with energies in the 10– 100 MeV range, at the dose of 2 Gy, for a 10 nm SV, $P(\epsilon > 0|D)$ is in the range from approximately 2.6×10^{-4} to 3.3×10^{-4} (Wang and Vassiliev 2014).

The Single Event Distribution of Deposited Energy, ϵ

The single event distribution $f_1(q)$ is the probability distribution of energy deposited in one energy deposition event. An example of an energy deposition event is a single charged particle entering the SV and depositing energy via one or more inelastic interactions that occur within the SV. Another example is two electrons produced by the same parent particle entering the same SV and both depositing energy. One reason why this should be considered a single event, instead of two separate events, is that the two particles reach the SV almost simultaneously, and a proportional counter registers this as a single event. Another reason is that we would prefer that $f_1(q)$ described interaction with the SV of the entire track produced by the source particle, including all daughter particles, as opposed to interactions with the SV of individual particles within this track.

The single event distribution is normalized

$$\int_0^\infty f_1(q)\,\mathrm{d}q = 1. \tag{6.2}$$

The point $q = 0$ is now included in the integral. This point, $q = 0$, is no longer singular and the function $f_1(q)$ is continuous as q tends to zero from the right.

The two distributions, $f(q|D)$ and $f_1(q)$, are related. At any given dose D, the number of energy deposition events is random and can be greater than one, or it can be zero. Given that we associate each event with the entire track from a single source particle, the distribution P_k of the number of energy deposition events is Poisson. For the same reason, we can assume that events are statistically independent, or equivalently, energy deposits ϵ_i from individual events are mutually independent random variables. We also note that the total energy ϵ deposited in all events is the sum of energies ϵ_i deposited in individual events, and that the distribution of a sum of random variables is the convolution of individual distributions of the summands (see Appendix A). It follows then that

$$f(q|D) = P_0\delta(q) + \sum_{k=1}^{\infty} P_k f_1^{*k}(q) = \sum_{k=0}^{\infty} \frac{\bar{N}^k}{k!}\mathrm{e}^{-\bar{N}} f_1^{*k}(q)\,, \tag{6.3}$$

where $\bar{N} = \bar{N}(D)$ is the average number of energy deposition events at dose D, $*k$ is the k-fold convolution, and $f_1^{*0}(q) \equiv \delta(q)$. In what follows, we will provide a simple formula for calculating $\bar{N}(D)$. Usually, the single event distribution, $f_1(q)$, is measured or calculated first, and then $f(q|D)$ is derived from the above equation.

The Specific Energy, z

The specific energy is the quotient of ϵ by the mass m of the material within the SV:

$$z = \frac{\epsilon}{m}. \tag{6.4}$$

The units of specific energy are J/kg, or Gy. It can be considered as a stochastic equivalent of the absorbed dose.

The Lineal Energy, y

The lineal energy is the quotient of the deposited energy ϵ by the mean chord length $\bar{l}$ of the SV:

$$y = \frac{\epsilon}{\bar{l}}. \tag{6.5}$$

The units of lineal energy are eV/nm, or keV/μm. It can be considered as a stochastic equivalent of the linear energy transfer (LET). The mean chord length in the above definition is not meant to represent the average actual path length of particles traversing the SV. It is a purely geometric concept. It is the mean chord length produced in the SV by uniformly and isotropically distributed infinite straight lines. This geometric model is a good approximation of a spatially uniform and isotropic fluence of heavy charged particles, if particle trajectories are almost straight, their ranges are much greater than the size of the SV, and, on the other hand, if the SV is sufficiently large to justify neglecting transport of delta electrons produced by the particles.

For this geometric model, the Cauchy theorem (Kellerer 1971) states that for a convex SV, the mean chord length is equal to

$$\bar{l} = \frac{4V}{S}, \tag{6.6}$$

where V is the volume of the SV, and S is its surface area. For example, for a spherical SV of radius R

$$\bar{l} = \frac{4}{3}R. \tag{6.7}$$

The actual, physical chord length of charged particles traversing the SV would be more meaningful in the context of radiobiological modeling, see for example Vassiliev (1995). The downside is that the physical chord length cannot be readily

measured, in any case not by the proportional counter. Still, the physical chord length is a useful quantity that can be calculated by the Monte Carlo method.

The Frequency Mean Lineal Energy, $\bar{y}_F$, and the Frequency Mean Specific Energy, $\bar{z}_F$

The frequency mean lineal and specific energies are the first moments of the single event distributions of lineal and specific energy, $f_1(y)$ and $f_1(z)$, respectively:

$$\bar{y}_F = \int_0^\infty y f_1(y)\,\mathrm{d}y. \tag{6.8}$$

$$\bar{z}_F = \int_0^\infty z f_1(z)\,\mathrm{d}z. \tag{6.9}$$

The average number of events, $\bar{N}(D)$ in Eq. (6.3), is

$$\bar{N}(D) = \frac{D}{\bar{z}_F}, \tag{6.10}$$

because D is the mean total energy from all energy deposition events divided by the mass of the SV, and $\bar{z}_F$ is the same but for one event.

The Dose-Weighted Single Event Distributions, $d(y)$ and $d(z)$

The dose-weighted distributions of lineal and specific energy are defined as

$$d(y) = \frac{y}{\bar{y}_F} f_1(y). \tag{6.11}$$

$$d(z) = \frac{z}{\bar{z}_F} f_1(z). \tag{6.12}$$

The Dose Mean Lineal Energy, $\bar{y}_D$, and the Dose Mean Specific Energy, $\bar{y}_D$

The dose weighted mean lineal and specific energies are the first moments of the dose weighted distributions of linear and specific energy, $d(y)$ and $d(z)$, respectively

$$\bar{y}_D = \int_0^\infty y\, d(y)\,\mathrm{d}y. \tag{6.13}$$

$$\bar{z}_D = \int_0^\infty z\, d(z)\,\mathrm{d}z. \tag{6.14}$$

6.3.2 *Application to Radiobiology*

The application of the microdosimetric approach to radiobiology is a very broad field. In this section, we give only a minimal introduction to the subject matter. For further reading, we refer to Rossi and Zaider (1996) and ICRU 36 (1983).

A general formalism for applying the microdosimetric approach to modeling biological effects is based on the following formula (Kellerer and Rossi 1972)

$$\text{Biological effect}\,(D) = \int_0^\infty B\,(z) f\,(z|D)\,\mathrm{d}z, \tag{6.15}$$

where $B(z)$ is the biological response function. It can be determined from a model of a biological effect, or it can be found by fitting the above formula to biological data for a set of different distributions $f(z|D)$.

The Theory of Dual Radiation Action (TDRA) by Kellerer and Rossi (1972) provides a classical example of the biological response function. The theory postulates that $B(z)$ is a quadratic function

$$B\,(z) = kz^2. \tag{6.16}$$

In another example (Zaider and Zaider 1985), the biological response function $B(z)$ was found by fitting a low dose version of Eq. (6.15) to experimental yields of chromosome aberrations in human lymphocytes for a range of radiation sources, from sparse radiations to several neutron sources and alpha emitters.

It follows from the above postulate of the TDRA that

$$\text{Biological effect}\,(D) = k\,(\bar{z}_D + D)\,D. \tag{6.17}$$

A derivation of the above result is given in Kellerer and Rossi (1972). Here, we will give a different proof, where we introduce a general technique for dealing with convolutions. It is based on the following property of the Laplace transform $\mathcal{L}$ (Arfken et al. 2013):

$$\mathcal{L}\,[f\,(z) * g\,(z)] = \mathcal{L}\,[f\,(z)]\,\mathcal{L}\,[g\,(z)]\,. \tag{6.18}$$

According to the TDRA, the biological effect is proportional to the second moment of distribution $f(z|D)$:

$$\text{Biological effect}\,(D) = \int_0^\infty kz^2 f\,(z|D)\,\mathrm{d}z = k\langle z^2\,(D)\rangle. \tag{6.19}$$

To calculate $\langle z^2\rangle$, we will use the following obvious formula:

$$\langle z^2\rangle = \left[\frac{\partial^2}{\partial^2 p}\langle \mathrm{e}^{-pz}\rangle\right]_{p=0}. \tag{6.20}$$

This means that instead of the average of z^2, we need to find the average of the exponent:

$$\langle \mathrm{e}^{-pz} \rangle = \int_0^\infty \mathrm{e}^{-pz} f\left(z|D\right) \mathrm{d}z = \mathrm{e}^{-\bar{N}} \sum_{k=0}^{\infty} \frac{\bar{N}^k}{k!} \int_0^\infty \mathrm{e}^{-pz} f_1^{*k}\left(z\right) \mathrm{d}z. \tag{6.21}$$

The last integral in the above equation is nothing but a Laplace transform of a k-fold convolution of the single event distribution $f_1(z)$. In that case, according to Eq. (6.18)

$$\int_0^\infty \mathrm{e}^{-pz} f_1^{*k}\left(z\right) \mathrm{d}z = \mathcal{L}\left[f_1\left(z\right)^{*k}\right] = f_1^k\left(p\right), \tag{6.22}$$

where $f_1(p)$ is the Laplace transform of $f_1(q)$. We now replace in Eq. (6.21) the integral by the right-hand side of Eq. (6.22). Then, we can see that the sum is the Taylor series expansion of an exponent:

$$\langle \mathrm{e}^{-pz} \rangle = \mathrm{e}^{-\bar{N}} \sum_{k=0}^{\infty} \frac{\bar{N}^k}{k!} f_1^k\left(p\right) = \mathrm{e}^{-\bar{N}} \mathrm{e}^{\bar{N} f_1(p)}. \tag{6.23}$$

The second derivative of the average of the exponent then can be easily found as

$$\frac{\partial^2}{\partial^2 p} \langle \mathrm{e}^{-pz} \rangle = \mathrm{e}^{-\bar{N}} \mathrm{e}^{\bar{N} f_1(p)} \left\{ \bar{N} \frac{\partial^2 f_1\left(p\right)}{\partial^2 p} + \left[\bar{N} \frac{\partial f_1\left(p\right)}{\partial p} \right]^2 \right\}. \tag{6.24}$$

To calculate the derivatives, we note that by the definition of the Laplace transform

$$f_1\left(p\right) = \int_0^\infty \mathrm{e}^{-pz} f_1\left(z\right) \mathrm{d}z. \tag{6.25}$$

From this, it can be seen that:

$$\left[f_1\left(p\right)\right]_{p=0} = 1. \tag{6.26}$$

$$\left[\frac{\partial f_1\left(p\right)}{\partial p} \right]_{p=0} = \int_0^\infty (-z) f_1\left(z\right) \mathrm{d}z = -\bar{z}_F. \tag{6.27}$$

$$\left[\frac{\partial^2 f_1\left(p\right)}{\partial^2 p} \right]_{p=0} = \int_0^\infty z^2 f_1\left(z\right) \mathrm{d}z \equiv \langle z^2 \rangle_F. \tag{6.28}$$

Finally, we set $p = 0$ in Eq. (6.24), and insert in the right-hand side Eqs. (6.26)–(6.28)

$$\langle z^2 \rangle = \left[\frac{\partial^2}{\partial^2 p} \langle \mathrm{e}^{-pz} \rangle \right]_{p=0} = \bar{N} \langle z^2 \rangle_F + \bar{N}^2 \left(-\bar{z}_F\right)^2 = D \bar{z}_D + D^2, \tag{6.29}$$

where in the last step we used Eq. (6.10), and

$$\langle z^2 \rangle_F = \bar{z}_D \bar{z}_F, \tag{6.30}$$

which follows from Eqs. (6.12) and (6.14). This proves Eq. (6.17). □

The dose mean specific energy $\bar{z}_D$, as well as the dose mean lineal energy $\bar{y}_D$, are often used as parameters of radiation quality. According to Eq. (6.17), in the low dose limit ($D \ll \bar{z}_D$) the relative biological effectiveness (RBE) is proportional to $\bar{z}_D$ or, alternatively, to $\bar{y}_D$. The latter parameter seems to be preferred in this context, because of its similarity to another popular parameter of radiation quality, the linear energy transfer (LET).

6.4 Intuitive Algorithms

The very small size of a typical SV poses serious challenges for Monte Carlo calculations. First, as we indicated earlier, a high spatial resolution of the track structure is required. For this reason, an event-by-event algorithm should be used for obtaining the best results. For micrometer sized SVs, accurate results can be achieved by using a carefully designed condensed history algorithm. In any case, algorithms achieving high spatial resolution are inherently slow. Another challenge associated with the small SV size is a small probability of a particle emitted by the source to reach the SV. In fact, the probability is so small that algorithms simply tracking particles from the source to the SV, without any variance reduction, are prohibitively slow, except for a few special cases. Making the problem even more difficult, standard variance reduction techniques generally cannot be used, because the Boltzmann equation does not apply to fluctuational characteristics of radiation fields, such as the microdosimetric spectrum, $f_1(q)$. For quantities of this nature, a stochastic transport equation must be used. An example of an equation of this type, and a Monte Carlo algorithm for solving it, are presented later in this section.

6.4.1 A Symmetry-Based Method

The most common approach to calculating microdosimetric characteristic by using the Monte Carlo method relies on the assumption of translational symmetry of the system on a scale exceeding the size of a particle track. It implies an infinite homogeneous medium, a spatially uniform source and fluence. The source could be spatially uniform in three dimensions, for example, when a spatially uniform fluence of photons is the source of delta electrons. The source could be uniform in a plane, representing, for example, a parallel beam of protons.

This method is most straightforward. It was described by Charlton et al. (1985), Nikjoo et al. (1989, 1997), and Palajova et al. (2006). In this method, first, a bounding volume V_B is determined, such that for a given particle energy it contains with a margin all particle tracks that start from the origin. The volume can be a cylinder for heavy charged particles (Charlton et al. 1985), or a sphere for electrons (Nikjoo et al. 1989). Then, a particle track is generated, and for all energy deposition events (inelastic collisions) the coordinates (x_i, y_i, z_i) of interaction points and the energies ϵ_i $(i = 1, 2, \cdots, N)$ deposited at the point of interaction are recorded. Next, a large number of SVs are placed randomly within V_B, and energies deposited in each SV are tallied. The procedure is repeated for a large number of tracks, until statistical uncertainties are reduced to the desired level.

For electrons, the algorithm becomes increasingly inefficient with increasing particle energy, because of an increasing track size and an increasing number of SVs with zero energy deposits. Algorithms have been developed for placing SVs closer to the track and for accelerating the search for interaction points that contribute to energy deposited in the SV (Nikjoo et al. 1991; Fitzsimons et al. 1998).

6.4.2 Sampling Over Individual Transfers

Another technique relying on translational symmetry is called sampling over individual transfers (Kellerer and Chmelevsky 1975). First, a particle track is generated in an infinite homogeneous medium and, similar to the method described above, coordinates of inelastic collisions (x_i, y_i, z_i) and respective energy deposits, ϵ_i, are recorded. Next, an energy deposition point i on this track is selected at random, with the selection probability proportional to the energy deposit at this point, ϵ_i. For the next step of the algorithm it is assumed that the SV is a sphere of radius R. In that case, next, the center of the SV is placed randomly within a sphere of radius R centered at the selected point i. Here, "randomly" means sampled from a uniform distribution. Then, the total energy ϵ deposited in all inelastic collisions within this randomly placed SV is tallied. Then another interaction point is selected randomly and a new location of the center of the SV is sampled, following the same procedure as above. The selection of an interaction point and placing of the SV are repeated a number of times. This number for any given track, should not be too large to avoid oversampling, which reduces the algorithm efficiency. Finally, all of the above steps, starting from generating the next track, are repeated until the uncertainties are reduced to the desired level. This method produces an estimate of the *dose weighted* distribution $d(z)$, because the algorithm for placing the SV favors locations of higher energy deposits, so that the probability of a location resulting in the total energy deposit ϵ is proportional to ϵ (Kellerer and Chmelevsky 1975).

The assumption of translational symmetry, common to the two methods, limits the category of problems that can be solved by using these methods. Another drawback is that generating and then analyzing entire tracks become increasingly time consuming with increasing particle energy.

6.4.3 Algorithm Based on Caswell's Analytical Method

Caswell (1966) developed an analytical method for calculating probability distributions of energy deposited in a spherical cavity by charged recoils, produced when a material is irradiated by neutrons. The method was formulated in the continuous slowing down approximation (CSDA). However, the underlying physical model of the method is very general and the method does not have to be limited to the CSDA transport model. It, in fact, can serve as a basis for a Monte Carlo algorithm.

In the Caswell method, all charged particles depositing energy in the SV are divided into two groups: particles that originated within the SV, and particles that originated outside the SV and entered it through its surface. Caswell classified all particles that originated within the SV either as "starters," if they exited the SV, or as "insiders," if they did not exit the SV. Similarly, particles that entered the SV through its surface and then exited were classified as "crossers," and those particles that stopped within the SV, as "stoppers." Because the SV is very small, it can be assumed that both the source function S and the fluence Φ are constant within the SV and on its surface, Γ. Thus, in this method, the small SV size is not an obstacle, on the contrary, it helps simplify the algorithm.

The average number of particles entering the SV is the product ΦA, where Φ is the total fluence and A is the cross-sectional area of the SV. This follows from definition 3 of fluence (Chap. 3). This, of course, is the fluence of charged particles. The number of particles originated within the SV is the product of the source function S and volume V of the SV. Here, S is the source function $S(x)$ integrated over energies and directions. The source function and fluence are not independent but rather, are related through the Boltzmann equation, $\hat{L}\Phi(x) = S(x)$. If the source function is normalized to one particle, the fluence must be normalized accordingly. If charged particles are produced via interactions of neutral particles, the source function is the collision density of neutral particles that accounts only for those collisions that produce charged particles. The energy distribution of "crossers" and "stoppers," when they enter the SV, is given by the fluence spectrum $\Phi(E)$. The distribution of the initial energy of "starters" and "insiders" is given by the source spectrum $S(E)$.

A Monte Carlo algorithm that follows from the above considerations can be summarized as follows. For now, we present it in a rather intuitive manner, without any further justification. We will return to this matter in the next subsection, after we introduce a more rigorous formalism.

Overview of the Algorithm

1. Randomly choose the particle's initial position within the volume V or on its surface Γ. The respective probabilities are proportional to $S \cdot V$ and $\Phi \cdot A$. As a side note, we mention that the ratio of the probabilities is $(\Phi \cdot A)/(S \cdot V)$. Recalling

that Φ is the total range of the particle, and that V/A is the mean chord length of the SV, we find that the ratio of the probabilities is proportional to the ratio of the particle track length to the mean chord length.

2. If the particle starts within V, sample its initial energy from the source spectrum, $S(E)$, and its initial location $\vec{r}_0$ uniformly within V. The latter distribution is uniform because we assumed above that the source function is constant within V. As for its initial direction $\vec{\Omega}_0$, in the case of a spherical SV, it can be sampled isotropically, or all particles can start in the same direction, for example, parallel to the z axis.
3. If the particle starts on Γ, sample its initial energy from the fluence spectrum $\Phi(E)$. Then, either sample its initial direction $\vec{\Omega}_0$ isotropically, or set it parallel to the z axis. Then, sample the initial location of the particle $\vec{r}_0$ on the surface Γ. The distribution of $\vec{r}_0$ on the surface Γ is not uniform. The initial point $\vec{r}_0$ is sampled on Γ so that the distribution of its projection in the direction $\vec{\Omega}_0$, on the cross-sectional area A of the SV, is uniform, where A is oriented normally to $\vec{\Omega}_0$. This distribution is identical to that of a parallel beam of protons incident on the SV. In the original formulation of this approach (Caswell 1966), only particles entering the SV, but not exiting, are considered at this step. In other words, if for a point $\vec{r}_0$ direction $\vec{\Omega}_0$ points away from the SV, the particle is discarded.
4. Generate a particle trajectory, and tally the energy deposited by the particle in the SV.

An advantage of this algorithm is that all particles reach the SV. Although, some particles may still traverse the SV without interaction, and thus deposit no energy. The downside is that the fluence spectrum $\Phi(E)$ at the SV location needs to be pre-calculated. In some important situations, this is a simple problem that requires solving only the energy degradation equation [see Sect. 3.9.3, Eq. (3.112)]. This is the case, for example, when the incident radiation consists of photons, and electronic equilibrium is established in the SV proximity. More generally, in this approach, the problem is separated into two problems that can be solved independently and, if appropriate, by using different numerical tools. The first problem is calculating the fluence spectrum $\Phi(E)$ at the SV location. This is a standard problem of radiation transport on a macroscopic scale, from a source to a detector. The second problem is calculating microdosimetric distributions from a source located in the SV and on its surface. This latter problem requires a high spatial resolution of the track structure, but only in the immediate vicinity of the SV.

Using the above algorithm with a more accurate transport model than the CSDA, for example an event-by-event Monte Carlo code, requires two corrections. First, we need to account for the possibility that a particle that left the SV, is scattered back towards the SV and reenters it. Second, we need to account for statistical correlations between particles entering the SV. These correlations exist because of the possibility of two or more daughter particles produced by the same primary particle entering the SV. The fluence spectrum $\Phi(E)$ provides no information on inter-particle correlations, and therefore in the above Monte Carlo algorithm particles starting at the SV surface have to be considered as independent. The assumption

of independence introduces errors into the calculated microdosimetric distribution $f_1(z)$ and its moments $\langle z^n \rangle$, for all n, except $n = 1$. In the next subsection, we will present a formalism that fully accounts for correlations and for reentering particles.

Alternatively, a simple modification of the above algorithm can be used to reduce these systematic errors. The SV can be placed at the center of a larger sphere. Then, all trajectories will start within the larger sphere or on its surface, in the same manner as described above. Of course, the energy that is deposited only in the SV, will be tallied. This modification of the algorithm is feasible because both particle correlations and the probability of back scattering into the SV decrease rapidly with increasing the distance from the SV. The downside of this modification is the reduced probability of a particle entering the SV.

6.5 Fluctuation Detector Method

The fluctuation detector method (FD) is based on the stochastic theory of particle transport. In the FD method, a form of the stochastic transport equation is solved by using the Monte Carlo method. The method was introduced in the TRION event-by-event code (Lappa et al. 1993). The FD method has two alternative implementations, FD-1 and FD-2 (Sects. 6.5.2 and 6.5.3).The algorithm presented in the preceding section is in fact a simplified version of the FD-2 implementation.

To derive the stochastic transport equation and formulate the algorithm, we use the Boltzmann equation and the adjoint transport equation for multiplying systems that were derived in Sects. 3.6 and 3.7. We have to consider multiplying systems, because correlations between particles entering the SV exist only in multiplying systems, where the particle emitted by the source produces secondary and higher generation particles.

6.5.1 Equation for the Distribution of Deposited Energy

We will now derive an equation for the distribution $f(q)$ of energy ϵ deposited in the SV. The distribution $f(q)$ is normalized to one source particle and satisfies

$$\int_0^\infty f\,(q)\,\mathrm{d}q = 1. \tag{6.31}$$

The moments of the distribution are defined as

$$\langle \epsilon^n \rangle = \int_0^\infty q^n f\,(q)\,\mathrm{d}q; \quad n = 1, 2 \ldots \tag{6.32}$$

As a matter of convenience, we define the zeroth moment so that the point $q = 0$ is not included in the integral:

$$\langle \epsilon^0 \rangle = \int_{0+}^{\infty} f(q)\,\mathrm{d}q. \tag{6.33}$$

This simply means that $\langle \epsilon^0 \rangle$ is the probability of an individual particle and/or its progenies depositing energy, $\epsilon > 0$, in the SV.

We start with an obvious equation

$$f(q) = \int S(x) f(q|x)\,\mathrm{d}x, \tag{6.34}$$

where $f(q|x)$ is the distribution of energy deposited in the SV by a particle that originated at the phase space point $x = (\vec{r}, \vec{\Omega}, E)$. To derive an equation for the conditional distribution $f(q|x)$, let us consider a particle with an initial phase coordinate x traveling a small distance l. With the probability σl the particle will undergo an interaction, and with the probability $1 - \sigma l$ it will travel the distance without interactions. Then,

$$f(q|x) = (1 - \sigma l) f(q|\, x,\ \text{no collisions}) + \sigma l f(q|\, x,\ \text{collision}). \tag{6.35}$$

If no collisions occur, then the particle simply moves from point $\vec{r}$ to point $\vec{r} + \vec{\Omega} l$:

$$f(q|\, x,\ \text{no collisions}) = f\left(q|\, \vec{r} + \vec{\Omega} l, \vec{\Omega}, E\right). \tag{6.36}$$

Then, we can combine the two terms in Eq. (6.35), and take the limit $l \to 0$:

$$\lim_{l \to 0} \frac{1}{l} \left[f(q|x) - (1 - \sigma l) f(q|\, x,\ \text{no collisions}) \right] = -\left(\vec{\Omega} \cdot \nabla\right) f(q|x) + \sigma f(q|x). \tag{6.37}$$

Similarly to Sects. 3.6 and 3.7 where the transport equations for a multiplying systems were introduced, here we consider a system that has only two types of interactions. The first type is scattering, during which the particle momentum changes from $\vec{p}$ to $\vec{p}\,'$, and no secondary particles are produced. The cross section of this process is $\sigma_1(\vec{p} \to \vec{p}\,')$. The second type is particle multiplication, during which the momentum of the incident particle changes, and a secondary particle of the same type as the incident particle is produced. The cross section of this process is $\sigma_2(\vec{p} \to \vec{p}\,', \vec{p}\,'')$, where $\vec{p}\,'$ and $\vec{p}\,''$ are the momenta of the two particles emerging from the collision. Then we can split the remaining term in Eq. (6.35) into two parts corresponding to the two types of collisions:

$$\sigma f\left(q|\, x \text{ collision}\right) = \int \mathrm{d}\vec{p}\,'\sigma_1\left(\vec{p} \to \vec{p}\,'\right) f\left(q - \epsilon_0^{(1)}|\vec{r},\vec{p}\,'\right) + \int \mathrm{d}\vec{p}\,' \int \mathrm{d}\vec{p}\,''\sigma_2\left(\vec{p} \to \vec{p}\,',\vec{p}\,''\right) f_2\left(q - \epsilon_0^{(2)}|\vec{r},\vec{p}\,',\vec{p}\,''\right). \tag{6.38}$$

If a collision occurs within the SV ($\vec{r} \in V$) we need to account for the energy deposited at $\vec{r}$ during the collision itself. This energy is the difference between the energy of the particle before the collision and the total energy of all particles emerging from it. Therefore, we introduce the two variables $\epsilon_0^{(1)}$ and $\epsilon_0^{(2)}$. These are equal to zero if the collision point $\vec{r}$ is outside the SV ($\vec{r} \notin V$), and otherwise ($\vec{r} \in V$) are equal to the energy deposited in a collision of type 1 and 2, respectively. Here, for brevity, we did not include particle absorption; another reason not to include it was because energy is deposited mostly by charged particles, for example electrons.

The two integrals in Eq. (6.38) are the total probability equations describing possible outcomes of the two types of collisions. Accordingly, the function $f_2\left(q|\vec{r},\vec{p}\,',\vec{p}\,''\right)$ is the probability distribution of the total energy ϵ deposited in the SV by two particles (and their progenies), that emerge with momenta $\vec{p}\,'$ and $\vec{p}\,''$ from a collision at point $\vec{r}$. The total energy deposit from two tracks is the sum of the energies deposited by the individual tracks: $\epsilon = \epsilon_1 + \epsilon_2$. The distribution of the sum is a convolution of individual distributions. Hence,

$$f_2\left(q|\vec{r},\vec{p}\,',\vec{p}\,''\right) = f\left(q|\vec{r},\vec{p}\,'\right) * f\left(q|\vec{r},\vec{p}\,''\right). \tag{6.39}$$

Taking the limit $l \to 0$ in Eq. (6.35) and using Eqs. (6.37)– (6.39) we obtain an equation for the conditional distribution $f(q|x)$. This equation is a form of the stochastic transport equation:

$$-\left(\vec{\Omega}\cdot\nabla\right) f\left(q|x\right) + \sigma f\left(q|x\right) = \int \mathrm{d}\vec{p}\,'\sigma_1\left(\vec{p} \to \vec{p}\,'\right) f\left(q - \epsilon_0^{(1)}|\vec{r},\vec{p}\,'\right) + \int \mathrm{d}\vec{p}\,' \int \mathrm{d}\vec{p}\,''\sigma_2\left(\vec{p} \to \vec{p}\,',\vec{p}\,''\right) f\left(q - \epsilon_0^{(2)}|\vec{r},\vec{p}\,'\right) * f\left(q - \epsilon_0^{(2)}|\vec{r},\vec{p}\,''\right). \tag{6.40}$$

The standard boundary condition for this equation is similar to the vacuum boundary condition for the adjoint function, that we specified in Sect. 3.4. If K is the surface defining the boundary of the computational domain and $\vec{n}_K$ is the outward pointing normal to K, the boundary condition for $f(q|x)$ is

$$f\left(q|\vec{r},\vec{p}\,\right) = \delta\left(q\right), \text{ if } \vec{r} \in K, \text{ and } \left(\vec{\Omega}\cdot\vec{n}_K\right) > 0. \tag{6.41}$$

It means that a particle starting on the surface of the computational domain, in an outward direction, $(\vec{\Omega}\cdot\vec{n}_K) > 0$, deposits zero energy in the SV.

6.5.2 Method FD-1

There are two versions of the FD method. Version FD-2 is equivalent to the algorithm based on the Caswell method, except that it is applicable to more accurate transport models than the CSDA. Version FD-1 is less intuitive, but is easier to implement because, in contrast to FD-2, it does not require the source spectrum, $S(E)$.

To derive the FD-1 algorithm, we first transform the integral in Eq. (6.34) as follows, where we use the inner product notation (f, g) for integrals over x:

$$\begin{aligned} f(q) &= (S(x), f(q|x)) = \left(\hat{L}\Phi(x), f(q|x)\right) \\ &= \left(\Phi(x), \hat{L}^{+} f(q|x)\right) \equiv (\Phi(x), F(q|x)). \end{aligned} \tag{6.42}$$

In the above equation, we first used the Boltzmann equation, $S = \hat{L}\Phi$, and then the definition of an adjoint operator (see Appendix B). In the last step, we introduced the notation $\hat{L}^{+} f \equiv F$. Comparing Eq. (6.42) with Eq. (3.28) that defines the detector response function $D(x)$, we observe that the function $F(q|x)$ can also be interpreted as a detector response function. This new detector "measures" fluctuations in radiation fields. The latter observation provided the name for the method.

A similar expression for the moments $\langle \epsilon^n \rangle$ of the distribution is derived by multiplying Eq. (6.42) by q^n ($n = 0, 1, 2, \ldots$) and then integrating it over q:

$$\langle \epsilon^n \rangle = \left(\Phi(x), F^{(n)}(x)\right), \tag{6.43}$$

where, obviously,

$$F^{(n)}(x) = \begin{cases} \int_0^\infty \mathrm{d}q\, q^n F(q|x), & n = 1, 2, \ldots; \\ \int_{0+}^\infty \mathrm{d}q\, F(q|x), & n = 0. \end{cases} \tag{6.44}$$

An important difference between detector response functions D and F is that the former is usually given, whereas the latter has to be calculated. The FD-1 method includes a Monte Carlo algorithm for calculating F. Although, in a simplified transport model, it may be possible to calculate F analytically. To derive an algorithm for calculating F, we have to use the explicit form of the adjoint operator $\hat{L}^{+}$. It is given by the left-hand side of the adjoint transport equation for multiplying systems, Eq. (3.67). Hence,

$$\begin{aligned} F(q|x) = \hat{L}^{+} f(q|x) = &- \left(\vec{\Omega} \cdot \nabla\right) f\left(q|\vec{r}, \vec{p}\right) + \sigma f\left(q|\vec{r}, \vec{p}\right) \\ &- \int \mathrm{d}\vec{p}\,' \sigma_1\left(\vec{p} \to \vec{p}\,'\right) f\left(q|\vec{r}, \vec{p}\,'\right) \\ &- \int \mathrm{d}\vec{p}\,' \int \mathrm{d}\vec{p}\,'' \sigma_2\left(\vec{p} \to \vec{p}\,', \vec{p}\,''\right) \left[f\left(q|\vec{r}, \vec{p}\,'\right) + f\left(q|\vec{r}, \vec{p}\,''\right)\right]. \end{aligned} \tag{6.45}$$

We notice that the first two terms on the right-hand side of Eq. (6.51) both are present in the stochastic equation for $f(q|x)$, Eq. (6.40). Then, we can replace these two terms with the integrals from the right-hand side of Eq. (6.40). This allows us to simplify the expression for F:

$$\begin{aligned} F(q|x) &= \int \mathrm{d}\vec{p}\,'\sigma_1\left(\vec{p}\to\vec{p}\,'\right) f\left(q-\epsilon_0^{(1)}|\vec{r},\vec{p}\,'\right) - \int \mathrm{d}\vec{p}\,'\sigma_1\left(\vec{p}\to\vec{p}\,'\right) f\left(q|\vec{r},\vec{p}\,'\right) \\ &+ \int \mathrm{d}\vec{p}\,' \int \mathrm{d}\vec{p}\,''\sigma_2\left(\vec{p}\to\vec{p}\,',\vec{p}\,''\right) f\left(q-\epsilon_0^{(2)}|\vec{r},\vec{p}\,'\right) * f\left(q-\epsilon_0^{(2)}|\vec{r},\vec{p}\,''\right) \\ &- \int \mathrm{d}\vec{p}\,' \int \mathrm{d}\vec{p}\,''\sigma_2\left(\vec{p}\to\vec{p}\,',\vec{p}\,''\right)\left[f\left(q|\vec{r},\vec{p}\,'\right)+f\left(q|\vec{r},\vec{p}\,''\right)\right] \\ &= \int \mathrm{d}\vec{p}\,'\sigma_1\left(\vec{p}\to\vec{p}\,'\right)\left[f\left(q-\epsilon_0^{(1)}|\vec{r},\vec{p}\,'\right)-f\left(q|\vec{r},\vec{p}\,'\right)\right] \\ &+ \int \mathrm{d}\vec{p}\,' \int \mathrm{d}\vec{p}\,''\sigma_2\left(\vec{p}\to\vec{p}\,',\vec{p}\,''\right) \\ &\times\left[f\left(q-\epsilon_0^{(2)}|\vec{r},\vec{p}\,'\right) * f\left(q-\epsilon_0^{(2)}|\vec{r},\vec{p}\,''\right) - f\left(q|\vec{r},\vec{p}\,'\right) - f\left(q|\vec{r},\vec{p}\,''\right)\right]. \end{aligned} \tag{6.46}$$

The convolution was introduced in Eq. (6.39) only for the sake of deriving a closed equation for the conditional distribution $f(q|x)$. Replacing the convolution with distribution f_2 results in an expression for F that has an obvious statistical interpretation that helps us design a Monte Carlo algorithm for calculating F and $F^{(n)}$.

$$\begin{aligned} F(q|x) &= \int \mathrm{d}\vec{p}\,'\sigma_1\left(\vec{p}\to\vec{p}\,'\right)\left[f\left(q-\epsilon_0^{(1)}|\vec{r},\vec{p}\,'\right)-f\left(q|\vec{r},\vec{p}\,'\right)\right] \\ &+ \int \mathrm{d}\vec{p}\,' \int \mathrm{d}\vec{p}\,''\sigma_2\left(\vec{p}\to\vec{p}\,',\vec{p}\,''\right) \\ &\times\left[f_2\left(q-\epsilon_0^{(2)}|\vec{r},\vec{p}\,',\vec{p}\,''\right) - f\left(q|\vec{r},\vec{p}\,'\right) - f\left(q|\vec{r},\vec{p}\,''\right)\right]. \end{aligned} \tag{6.47}$$

To derive a formula for the functions $F^{(n)}(x)$, we multiply Eq. (6.47) by q^n and then integrate it over q. To make the statistical interpretation of the result more obvious, we will express it in terms of conditional expectations. To do so, we introduce random variables ϵ_1 and ϵ_2 that denote the energy deposited in the SV by particles starting from point $\vec{r}$ with momenta $\vec{p}\,'$ and $\vec{p}\,''$, respectively. Then,

$$\begin{aligned} F^{(n)}(x) &= \int \mathrm{d}\vec{p}\,'\sigma_1\left(\vec{p}\to\vec{p}\,'\right)\left[\mathrm{E}\left\{\left(\epsilon_0^{(1)}+\epsilon_1\right)^n|\vec{r},\vec{p}\,'\right\} - \mathrm{E}\left\{\epsilon_1^n|\vec{r},\vec{p}\,'\right\}\right] \\ &+ \int \mathrm{d}\vec{p}\,' \int \mathrm{d}\vec{p}\,''\sigma_2\left(\vec{p}\to\vec{p}\,',\vec{p}\,''\right) \\ &\times\left[\mathrm{E}\left\{\left(\epsilon_0^{(2)}+\epsilon_1+\epsilon_2\right)^n|\vec{r},\vec{p}\,',\vec{p}\,''\right\} - \mathrm{E}\left\{\epsilon_1^n|\vec{r},\vec{p}\,'\right\} - \mathrm{E}\left\{\epsilon_2^n|\vec{r},\vec{p}\,''\right\}\right], \end{aligned} \tag{6.48}$$

where $\mathrm{E}\{\xi|A\}$ denotes the expectation of ξ under condition A.

To make the algorithm even more obvious, we note that the differential cross sections $\sigma_1(\vec{p} \to \vec{p}\,')$ and $\sigma_2(\vec{p} \to \vec{p}\,', \vec{p}\,'')$ can be expressed in terms of the total cross sections $\sigma_1(\vec{p}\,)$, $\sigma_2(\vec{p}\,)$, $\sigma(\vec{p}\,) = \sigma_1(\vec{p}\,) + \sigma_1(\vec{p}\,)$, and the respective distribution functions of the momenta, $g_1(\vec{p}\,'|\vec{p}\,)$ and $g_2(\vec{p}\,', \vec{p}\,''|\vec{p}\,)$:

$$\sigma_1\left(\vec{p} \to \vec{p}\,'\right) = \sigma\left(\vec{p}\,\right) \frac{\sigma_1\left(\vec{p}\,\right)}{\sigma\left(\vec{p}\,\right)} g_1(\vec{p}\,'|\vec{p}\,). \tag{6.49}$$

$$\sigma_2\left(\vec{p} \to \vec{p}\,', \vec{p}\,''\right) = \sigma\left(\vec{p}\,\right) \frac{\sigma_2\left(\vec{p}\,\right)}{\sigma\left(\vec{p}\,\right)} g_2(\vec{p}\,', \vec{p}\,''|\vec{p}\,). \tag{6.50}$$

Later in this section, we will make additional approximations and present a more practical version of the algorithm. At this point we give only an overview of the algorithm in the most general form and, for simplicity, we do so only for $F^{(n)}(x)$.

Algorithm Overview

1. Simulate an interaction at point $\vec{r}$ of an incident particle with momentum $\vec{p}$. The interaction type 1 or 2 is chosen at random. The probability of type 1 is $\sigma_1(\vec{p}\,)/\sigma(\vec{p}\,)$ and that of type 2 is $\sigma_2(\vec{p}\,)/\sigma(\vec{p}\,)$.
2. If $\vec{r} \in V$, determine $\epsilon_0^{(1)}$ or $\epsilon_0^{(2)}$, depending on the type of the interaction. If $\vec{r} \notin V$, set $\epsilon_0^{(1)} = \epsilon_0^{(2)} = 0$.
3. Generate the trajectories of all particles produced in the interaction and their progenies.
4. If only one particle was produced in the collision, tally the energy ϵ_1 its entire track deposits in the SV. If two particles were produced, then separately tally the energies ϵ_1 and ϵ_2 that are deposited in the SV by each of the two tracks.
5. If the interaction type was 1, calculate

$$\xi = \left(\epsilon_0^{(1)} + \epsilon_1\right)^n - \epsilon_1^n. \tag{6.51}$$

   Otherwise, calculate

$$\xi = \left(\epsilon_0^{(2)} + \epsilon_1 + \epsilon_2\right)^n - \epsilon_1^n - \epsilon_2^n. \tag{6.52}$$

6. Repeat steps 1–5 many times. Generate a sample $\{\xi_1, \xi_2, \ldots \xi_N\}$, $N \gg 1$.
7. Estimate the detector response function

$$F^{(n)}(x) \approx \frac{\sigma\left(\vec{p}\,\right)}{N} \sum_{i=1}^{N} \xi_i. \tag{6.53}$$

From Eqs. (6.51) and (6.52) two important observations can be made.

1. If $\vec{r} \notin V$, then $\xi = 0$ in Eq. (6.51). This means that for points outside the SV, there is no need to simulate interactions of type 1.
2. If $\vec{r} \notin V$, and two particles were produced in the interaction, but only one reached the SV, that is $\epsilon_1 > 0$ and $\epsilon_0^{(2)} = \epsilon_2 = 0$. In this case, in Eq. (6.52), we have $\xi = 0$.

These two properties are crucial for the feasibility of the FD method. If point $\vec{r}$ is outside the SV, then the estimator ξ is not zero only when the tracks of both particles reach the SV and deposit energy, $\epsilon_1 > 0$ and $\epsilon_2 > 0$. The probability of this happening decreases rapidly with increasing the distance from the SV. One reason for this is purely geometrical. Given the very small size of a typical SV, the probability that even one particle reaches it rapidly diminishes with distance. Another reason is the finite range of charged particles. The latter is particularly important for delta electrons that tend to have a low energy. All of the above implies that $F^{(n)}$ decreases rapidly with increasing the distance from the SV.

Then, a distance $\Delta R_{\max}$ from the surface of the SV can be found, such that setting $F^{(n)}(x) = 0$ for points beyond $\Delta R_{\max}$ does not introduce significant errors. In this case, it is convenient to separate the integration over $\vec{r}$ in Eq. (6.43) into an integral over V (the SV) and an extension volume $\tilde{V}$:

$$\langle \epsilon^n \rangle = \left(\Phi(x), F^{(n)}(x) \right)_V + \left(\Phi(x), F^{(n)}(x) \right)_{\tilde{V}}, \tag{6.54}$$

Assuming that the SV is a sphere, $\tilde{V}$ is a spherical shell of thickness $\Delta R_{\max}$ around it. The two integrals can be calculated by using the Monte Carlo method, by first sampling a point $\vec{r}$ in V or $\tilde{V}$, respectively, and then calculating $F^{(n)}(x)$ following the algorithm we outlined above. In this case generating only one ξ for each sampled $\vec{r}$ will suffice. The algorithm can be simplified further, if the SV and the extension volume are sufficiently small, so that the fluence can be approximated by a constant in the volume $V \cup \tilde{V}$. In this case,

$$\langle \epsilon^n \rangle = \int \mathrm{d}E \int \mathrm{d}\vec{\Omega}\, \Phi\left(\vec{\Omega}, E\right) \left[\int_V \mathrm{d}\vec{r}\, F^{(n)}\left(\vec{r}, \vec{\Omega}, E\right) + \int_{\tilde{V}} \mathrm{d}\vec{r}\, F^{(n)}\left(\vec{r}, \vec{\Omega}, E\right) \right]. \tag{6.55}$$

For a spherical SV and a spherical shell $\tilde{V}$, the two integrals in the square brackets are independent of the direction $\vec{\Omega}$. Then, the result simplifies even further:

$$\langle \epsilon^n \rangle = \int \mathrm{d}E\, \Phi(E) \left[\int_V \mathrm{d}\vec{r}\, F^{(n)}\left(\vec{r}, E\right) + \int_{\tilde{V}} \mathrm{d}\vec{r}\, F^{(n)}\left(\vec{r}, E\right) \right], \tag{6.56}$$

where $\Phi(E)$ is the fluence spectrum in $V \cup \tilde{V}$, and functions $F^{(n)}$ are calculated for an arbitrary direction $\vec{\Omega}$. Finally, by using the following notation

$$G^{(n)}(E) = \int_V \mathrm{d}\vec{r}\, F^{(n)}\left(\vec{r}, E\right) + \int_{\tilde{V}} \mathrm{d}\vec{r}\, F^{(n)}\left(\vec{r}, E\right) \equiv G_V^{(n)}(E) + G_{\tilde{V}}^{(n)}(E), \tag{6.57}$$

the moments of deposited energy can be written as a one-dimensional integral:

$$\langle \epsilon^n \rangle = \int \Phi(E)\, G^{(n)}(E)\, \mathrm{d}E \tag{6.58}$$

The functions $G^{(n)}(E)$ depend on the particle type and the size of the SV. They can be conveniently tabulated. If $G^{(n)}(E)$ are pretabulated, the calculation of fluctuational characteristics of energy deposition is not different from the standard problem of calculating linear functionals of a solution of the Boltzmann equation. This simplicity is, perhaps, misleading, because we actually solve a form of the stochastic transport equation, Eq. (6.40). The overall strategy of the FD method consists of breaking the problem in two parts. The first part is calculation of the detector response functions $G^{(n)}(E)$. This part involves radiation transport only in the vicinity of the SV and, therefore, it can be completed with an event-by-event algorithm, despite the slow speed of algorithms of this type. The second part is calculating the fluence at the SV location and its functional given by Eq. (6.58). This is the standard problem of radiation transport, for which all the powerful methods and optimization techniques for solving the Boltzmann equation can be employed. Under the condition of electronic equilibrium, the fluence spectrum of electrons can be found by solving the energy degradation equation [Sect. 3.9.3, Eq. (3.112)]. The latter is a one-dimensional integral equation that can be solved fast by using the Monte Carlo algorithm that was described in Sect. 4.3.5.

We can now present an algorithm for calculating $G^{(n)}(E)$. We start with $G_V^{(n)}(E)$.

Algorithm Groundwork

Determine the size of the computational domain. Its boundary K should be placed at such a distance from the SV that particles reaching it would have a negligible probability of scattering back and entering the SV.

Algorithm

1. Place a point $\vec{r}$ at a random location within the SV. The distribution of $\vec{r}$ is uniform.
2. Randomly choose the interaction type. The probability of choosing a type 1 interaction is $\sigma_1(E)/\sigma(E)$ and that of choosing type 2 interaction is $\sigma_2(E)/\sigma(E)$.
3. Simulate the chosen interaction type, at point $\vec{r}$, of an incident particle with energy E traveling parallel to the z axis in the positive direction.
4. Determine the energy deposited in the SV from the interaction itself, $\epsilon_0^{(1)}$ or $\epsilon_0^{(2)}$, depending on the type of the interaction.

5. Generate trajectories of all particles produced in the interaction and their progenies. Particle trajectories are terminated when they reach the boundary K of the computational domain or when their energy falls below the tracking cut-off.
6. If only one particle was produced in the collision, tally the energy ϵ_1 its entire track deposits in the SV. If two particles were produced, then separately tally the energies ϵ_1 and ϵ_2 deposited in the SV by each of the two tracks.
7. If the interaction was type 1, calculate

$$\xi = \left(\epsilon_0^{(1)} + \epsilon_1\right)^n - \epsilon_1^n. \tag{6.59}$$

Otherwise, calculate

$$\xi = \left(\epsilon_0^{(2)} + \epsilon_1 + \epsilon_2\right)^n - \epsilon_1^n - \epsilon_2^n. \tag{6.60}$$

8. Repeat steps 1–7 many times. Generate a sample $\{\xi_1, \xi_2, \ldots \xi_N\}$, $N \gg 1$.
9. Estimate the detector response function as

$$G_V^{(n)}(E) \approx \sigma(E) \frac{V}{N} \sum_{i=1}^{N} \xi_i. \tag{6.61}$$

Calculation of function the $G_{\tilde{V}}^{(n)}(E)$ follows the same steps, except:

1. The point of interaction $\vec{r}$ is placed in the extension volume, $\tilde{V}$.
2. Only type 2 interactions are simulated.
3. Accordingly, Eq,(6.59) is not used.
4. In Eq. (6.61) the SV volume is replaced by the extension volume:

$$G_{\tilde{V}}^{(n)}(E) \approx \sigma(E) \frac{\tilde{V}}{N} \sum_{i=1}^{N} \xi_i. \tag{6.62}$$

Calculation of the distribution of deposited energy, $f(q)$, follows the same steps as calculation of the moments $\langle \epsilon^n \rangle$, except that Eqs. (6.58)–(6.61) have to be modified, as described below. First, Eq. (6.58) is replaced by

$$f(q) = \int \Phi(E) G(q|E) \, \mathrm{d}E. \tag{6.63}$$

Further, we have to approximate the distribution $f(q)$ by a histogram f_j ($j = 1, 2, \ldots, k$) and, accordingly, replace $G(q|E)$ with $G(j, E)$. The latter is, again, calculated as a sum $G(j, E) = G_V(j, E) + G_{\tilde{V}}(j, E)$. Then, we need to define the boundaries of the energy deposition bins: $q_0, q_1, \ldots q_k$, bin widths $\Delta q_j = q_j - q_{j-1}$,

and an indicator function $I_j(\epsilon)$ that returns 1 if ϵ belongs to the energy bin j, and returns 0, otherwise. Then, to calculate $G_V(j,E)$ we replace Eqs. (6.59)–(6.61) with the following:

$$\xi_j = I_j\left(\epsilon_0^{(1)} + \epsilon_1\right) - I_j\left(\epsilon_1\right). \tag{6.64}$$

$$\xi_j = I_j\left(\epsilon_0^{(2)} + \epsilon_1 + \epsilon_2\right) - I_j\left(\epsilon_1\right) - I_j\left(\epsilon_2\right). \tag{6.65}$$

$$G_V\left(j,E\right) \approx \frac{\sigma\left(E\right)V}{N\Delta q_j}\sum_{i=1}^{N}\left(\xi_j\right)_i;\ j=1,2,\ldots,k. \tag{6.66}$$

In the corresponding formula for $G_{\tilde{V}}(j,E)$, volume V on the right-hand side of Eq. (6.66) is replaced by $\tilde{V}$.

6.5.3 Method FD-2

In Eq. (6.34) the distribution of deposited energy $f(q)$ is written as an integral over the phase space. We split this integral into two integrals, one over the volume V of the SV, and the other over the complement volume $\bar{V}$:

$$f\left(q\right) = \left(S\left(x\right), f\left(q|x\right)\right)_V + \left(S\left(x\right), f\left(q|x\right)\right)_{\bar{V}}. \tag{6.67}$$

This equation is equivalent to Eq. (3.53), except that instead of the adjoint function Φ^+, we now have the conditional distribution $f(q|x)$. Next, we perform the same transformation of the above formula, as the one that was performed in Sect. 3.4. We leave the integral over V as is, and transform the integral over $\bar{V}$ into an expression similar to $(\Phi(x), F(q|x))$. The algebra is exactly the same as that in Sect. 3.4, so we can omit the details and proceed directly to the result

$$\begin{aligned} f\left(q\right) = &\left(S\left(x\right), f\left(q|x\right)\right)_V + \left(\Phi\left(x\right), F\left(q|x\right)\right)_{\bar{V}} \\ &+ \int \mathrm{d}\vec{\Omega}\int \mathrm{d}E \oint_{\Gamma(\bar{V})} \mathrm{d}\Gamma\ \left(\vec{\Omega}\cdot\vec{n}_\Gamma\right)\Phi\left(\vec{r},\vec{\Omega},E\right) f\left(q|\vec{r},\vec{\Omega},E\right). \end{aligned} \tag{6.68}$$

This equation is the same as Eq. (3.59), except that it is written in terms of the fluctuational characteristics, $f(q|x)$ and $F(q|x)$ instead of the adjoint function $\Phi^+(x)$ and the detector response function $D(x)$. Next, we simplify the three terms in this equation, one at a time. We assume that the SV is a sphere. To simplify the first term, we further assume that the source function is constant in V:

$$f_S\left(q\right) \equiv \left(S\left(x\right), f\left(q|x\right)\right)_V = \int \mathrm{d}\vec{\Omega}\int \mathrm{d}E\, S\left(\vec{\Omega},E\right)\int_V \mathrm{d}\vec{r}\, f\left(q|\vec{r},\vec{\Omega},E\right). \tag{6.69}$$

Owing to the spherical symmetry, the last integral is the same for all directions, $\vec{\Omega}$. Then,

$$f_S(q) = \int \mathrm{d}E\, S(E) \int_V \mathrm{d}\vec{r}\, f\left(q|\vec{r}, \vec{\Omega}_0, E\right) \equiv \int \mathrm{d}E\, S(E)\, G_S(q|E)\,, \tag{6.70}$$

where $S(E)$ is the source spectrum in V, $\vec{\Omega}_0$ is an arbitrary direction, for example, the direction of the z-axis, and G_S is the response function corresponding to the source term. For the distribution moments, we have

$$\langle \epsilon^n \rangle_S = \int \mathrm{d}E\, S(E) \int_V \mathrm{d}\vec{r}\, \mathrm{E}\left\{\epsilon^n|\vec{r}, \vec{\Omega}_0, E\right\} \equiv \int \mathrm{d}E\, S(E)\, G_S^{(n)}(E)\,, \tag{6.71}$$

Function $G_S^{(n)}(E)$ can be calculated by using Monte Carlo method. The algorithm is described in what follows.

Algorithm

1. Place a point $\vec{r}$ at a random location in V. The distribution of $\vec{r}$ is uniform.
2. Generate a track of a particle that starts at the point $\vec{r}$ with energy E, in the direction $\vec{\Omega}_0$, which is the positive direction of the z axis.
3. Tally the energy ϵ this particle and all its progenies deposit in the SV.
4. Repeat steps 1–3 many times. Generate a sample $\{\epsilon_1, \epsilon_2, \ldots, \epsilon_N, \}$, $N \gg 1$.
5. Estimate the response function

$$G_S^{(n)}(E) \approx \frac{V}{N} \sum_{i=1}^{N} \epsilon_i^n. \tag{6.72}$$

In the second term in Eq. (6.68), we can replace the complement volume $\bar{V}$ with the extension volume $\tilde{V}$, because, as we have discussed, $F(q|x)$ decreases rapidly with increasing the distance from the SV and becomes negligible outside $\tilde{V}$:

$$f_{\bar{V}}(q) \equiv (\Phi(x), F(q|x))_{\bar{V}} \approx (\Phi(x), F(q|x))_{\tilde{V}}\,. \tag{6.73}$$

This term was considered in the FD-1 method, and the algorithm has already been presented. This term is calculated using particle trajectories that originate outside the SV.

Finally, the last term in Eq. (6.68) involves integration over the surface Γ of the complement volume $\bar{V}$. This surface, of course, coincides with the surface of the SV. The outward pointing normal to the surface $\Gamma(\bar{V})$, $\vec{n}_\Gamma$, points into the SV, at its center. This means that for particles entering the SV, we have $(\vec{\Omega} \cdot \vec{n}_\Gamma) > 0$, and for particles leaving the SV, we have $(\vec{\Omega} \cdot \vec{n}_\Gamma) < 0$.

Again, we assume that the SV is sufficiently small, so that the fluence can be approximated by a constant, this time for all points on the surface Γ. This allows to simplify the integral, first by removing the fluence from the surface integral

$$f_\Gamma(q) \equiv \int \mathrm{d}\vec{\Omega} \int \mathrm{d}E \;\; \Phi\left(\vec{\Omega}, E\right) \oint_{\Gamma(\bar{V})} \mathrm{d}\Gamma \;\; \left(\vec{\Omega} \cdot \vec{n}_\Gamma\right) f\left(q|\vec{r}, \vec{\Omega}, E\right), \tag{6.74}$$

and then separating the surface integral into two parts, corresponding to the vector $\vec{\Omega}$ pointing in and out of the SV.

$$\begin{aligned} f_\Gamma(q) = & \int \mathrm{d}\vec{\Omega} \int \mathrm{d}E \;\; \Phi\left(\vec{\Omega}, E\right) \\ & \times \left[\oint_{\text{in}} \mathrm{d}\Gamma \;\; |\vec{\Omega} \cdot \vec{n}_\Gamma| f\left(q|\vec{r}, \vec{\Omega}, E\right) - \oint_{\text{out}} \mathrm{d}\Gamma \;\; |\vec{\Omega} \cdot \vec{n}_\Gamma| f\left(q|\vec{r}, \vec{\Omega}, E\right) \right]. \end{aligned} \tag{6.75}$$

Here, $\Phi(\vec{\Omega}, E)$ is the fluence on the surface Γ. Owing to the spherical symmetry, each of the two surface integrals is independent of $\vec{\Omega}$. It follows then that

$$\begin{aligned} f_\Gamma(q) = & \int \mathrm{d}E \;\; \Phi(E) \\ & \times \left[\oint_{\text{in}} \mathrm{d}\Gamma \; |\vec{\Omega}_0 \cdot \vec{n}_\Gamma| f\left(q|\vec{r}, \vec{\Omega}_0, E\right) - \oint_{\text{out}} \mathrm{d}\Gamma \;\; |\vec{\Omega}_0 \cdot \vec{n}_\Gamma| f\left(q|\vec{r}, \vec{\Omega}_0, E\right) \right] \\ & \equiv \int \mathrm{d}E \;\; \Phi(E) \left[G_{\text{in}}(q|E) - G_{\text{out}}(q|E) \right], \end{aligned} \tag{6.76}$$

where $\vec{\Omega}_0$ is an arbitrarily chosen direction, and G_{in} and G_{out} are the response functions. Similarly, for the moments of distribution, we have

$$\begin{aligned} \langle \epsilon^n \rangle_\Gamma = & \int \mathrm{d}E \; \Phi(E) \\ & \times \left[\oint_{\text{in}} \mathrm{d}\Gamma \; |\vec{\Omega}_0 \cdot \vec{n}_\Gamma| \mathrm{E}\left\{ \epsilon^n | \vec{r}, \vec{\Omega}_0, E \right\} - \oint_{\text{out}} \mathrm{d}\Gamma \; |\vec{\Omega}_0 \cdot \vec{n}_\Gamma| \mathrm{E}\left\{ \epsilon^n | \vec{r}, \vec{\Omega}_0, E \right\} \right] \\ & \equiv \int \mathrm{d}E \; \Phi(E) \left[G_{\text{in}}^{(n)}(E) - G_{\text{out}}^{(n)}(E) \right]. \end{aligned} \tag{6.77}$$

The response functions $G_{\text{in}}(q|E)$, $G_{\text{out}}(q|E)$, $G_{\text{in}}^{(n)}(E)$ and $G_{\text{out}}^{(n)}(E)$ can be calculated by using the Monte Carlo method. Before presenting the algorithm, we note that $\mathrm{d}\Gamma|\vec{\Omega}_0 \cdot \vec{n}_\Gamma|$ is the projection of the surface area $\mathrm{d}\Gamma$ onto a plane normal to the direction $\vec{\Omega}_0$. Therefore,

$$\oint_{\Gamma(\bar{V})} \mathrm{d}\Gamma \, |\vec{\Omega}_0 \cdot \vec{n}_\Gamma| = A, \tag{6.78}$$

where A is the cross-sectional area of the SV. Furthermore, to calculate the surface integrals in Eq. (6.77) we need to place a point $\vec{r}$ randomly on the surface Γ. If we do so using a uniform over the surface distribution, in this case over a half of the spherical surface (when calculating either the "in" or "out" integral), then the expectation values will have to be weighted by $|\vec{\Omega}_0 \cdot \vec{n}_\Gamma(\vec{r}\,)|$, where $\vec{n}_\Gamma(\vec{r}\,)$ is the normal to the surface at point $\vec{r}$. The weight is not needed, if we sample point $\vec{r}$ from a distribution proportional to $|\vec{\Omega}_0 \cdot \vec{n}_\Gamma|$. This distribution is such that the projection of a point $\vec{r}$ on the cross-sectional area normal to $\vec{\Omega}_0$ is uniformly distributed over the cross-sectional area. In other words, this distribution is the same as that of a parallel beam of particles traveling in the direction $\vec{\Omega}_0$.

We now present an algorithm for calculating $G^{(n)}_{\mathrm{in}}(E)$. Algorithms for the other three response functions are similar. We place the origin of the coordinate system at the center of the SV, and choose the direction $\vec{\Omega}_0$ to be the same as the positive direction of the z-axis. We also need to define the boundary K of the computational domain.

Algorithm

1. Place a point $\vec{r}$ at a random location on that part of the SV surface Γ, for which $z < 0$. This part is a hemisphere. For this hemisphere, the direction $\vec{\Omega}_0$ is "in." A particle starting from this hemisphere enters the SV. The distribution of $\vec{r}$ is proportional to $|\vec{\Omega}_0 \cdot \vec{n}_\Gamma(\vec{r}\,)|$.
2. Generate a track of a particle that starts from point $\vec{r}$, in the direction $\vec{\Omega}_0$, with energy E. Terminate the particle trajectory when it leaves the computational domain.
3. Tally the energy ϵ this particle and all its progenies deposit in the SV.
4. Repeat steps 1–3 many times. Generate a sample $\{\epsilon_1, \epsilon_2, \ldots, \epsilon_N\}$, $N \gg 1$.
5. Estimate the response function

$$G^{(n)}_{\mathrm{in}}(E) \approx \frac{A}{N} \sum_{i=1}^{N} \epsilon_i^n. \tag{6.79}$$

The algorithm for calculating $G^{(n)}_{\mathrm{out}}(E)$ is the same, except particle trajectories start on the other hemisphere, for which $z > 0$. These particles start in the direction away from the SV. Algorithms for calculating $G_{\mathrm{in}}(q|E)$ and $G_{\mathrm{out}}(q|E)$ are different from the above only in that the sample $\{\epsilon_1, \epsilon_2, \ldots, \epsilon_N\}$ is used for estimating the distributions of ϵ, that is, for generating a histogram, instead of estimating the momenta.

The FD-2 method is indeed very similar to the algorithm based on the Caswell approach (Sect. 6.4.3). There are, however, differences. The FD-2 method has two terms that are not included in the Caswell method. These terms are $f_{\bar{V}}$, Eq. (6.73), and G_{out}, Eqs. (6.76)–(6.77). The first of the two terms corrects the result for possible correlations between particles entering the SV. As we discussed earlier, to calculate this term, we start trajectories from collisions that produce at least two particles, and it is nonzero only when both particles directly, or through their progenies, deposit energy in the SV. The second term G_{out} accounts for contribution from particles that backscattered into the SV after they leave it. These two terms typically are small corrections, and they were not needed in the original formulation of the Caswell method that used the CSDA transport model. However, when a more accurate transport model is used, such as the event-by-event Monte Carlo algorithm, these corrections are no longer negligible.

6.6 The Effect of Energy Straggling on Microdosimetric Spectra

Exercise

Consider a parallel beam of hypothetical heavy charged particles incident on a small spherical SV. The particles travel along straight lines without deflections. The mean free path between inelastic interactions is 10 nm. In each interaction, the particles deposit 10 eV at the point of interaction. The particle properties do not change as it loses energy, nor does it produce any secondary particles.

1. Calculate the stopping power of these particles. Compare it with the stopping power of protons in the tissue. Estimate the proton energy.
2. Calculate analytically the microdosimetric spectrum $f_1(q)$. Use the CSDA approximation, assuming a constant stopping power. The chord length distribution is given in Sect. 2.2, Eq. (2.3).
3. Write a Monte Carlo event-by-event code based on the described transport model.

Algorithm Overview

We assume that the SV is a sphere centered at the origin, its radius is R, and the particles travel in the positive z direction.

(a) Start a particle trajectory at $z = -R$, just outside the sphere. Sample its initial x and y as coordinates of a point uniformly distributed within a circle of radius R.

(b) Generate a particle trajectory. The free path is sampled from the exponential distribution, Eq. (2.7). The trajectory is terminated when the particles leaves the SV.
(c) Tally the energy ϵ the particle deposits in the SV. The contribution to ϵ from each inelastic collision within the SV is 10 eV.
(d) Repeat steps (a)–(c) many times. Generate a sample $\{\epsilon_1, \epsilon_2, \ldots, \epsilon_N, \}$, $N \gg 1$.
(e) Use the sample to make a histogram (see Sect. 2.14.1). Compare the histogram with the analytical distribution $f_1(q)$.

4. Perform the above calculations for R=100, 1000, and 10,000 nm.

References

Arfken, G.B., Weber, H.J., Harris, F.E.: Mathematical Methods for Physicists, 7th edn. Elsevier, Amsterdam (2013)

Bell, G.I.: On stochastic theory of neutron transport. Nucl. Sci. Eng. **21**(3), 390–401 (1965)

Bernal, M.A., Bordage, M.C., Brown, J.M.C., Davidkova, M., Delage, E., El Bitar, Z., Enger, S.A., Francis, Z., Guatelli, S., Ivanchenko, V.N., Karamitros, M., Kyriakou, I., Maigne, L., Meylan, S., Murakami, K., Okada, S., Payno, H., Perrot, Y., Petrovic, I., Pham, Q.T., Ristic-Fira, A., Sasaki, T., Stepan, V., Tran, H.N., Villagrasa, C., Incerti, S.: Track structure modeling in liquid water: a review of the Geant4-DNA very low energy extension of the Geant4 Monte Carlo simulation toolkit. Phys. Medica **31**(8), 861–874 (2015)

Caswell, R.S.: Deposition of energy by neutrons in spherical cavities. Radiat. Res. **27**(8), 92–107 (1966)

Charlton, D.E., Goodhead, D.T., Wilson, W.E., Paretzke, H.G.: Energy deposition in cylindrical volumes: a) protons energy 0.3 MeV to 4.0 MeV b) alpha particles energy 1.0 MeV to 20.0 MeV. Monograph 8 5/1 Chilton: MRC Radiobiology Unit (1985)

Fitzsimons, C.J., Nikjoo, H., Bolton, C.E., Goodhead, D.T.: A novel algorithm for tracing the interaction of a track with molecular targets - use of Delaunay triangulation. Math. Biosci. **154**(2), 103–115 (1998)

ICRU: Microdosimetry. ICRU Report 36, Bethesda, MD (1983)

Incerti, S., Ivanchenko, A., Karamitros, M., Mantero, A., Moretto, P., Tran, H.N., Mascialino, B., Champion, C., Ivanchenko, V.N., Bernal, M.A., Francis, Z., Villagrasa, C., Baldacchino, G., Gueye, P., Capra, R., Nieminen, P., Zacharatou, C.: Comparison of Geant4 very low energy cross section models with experimental data in water. Med. Phys. **37**(9), 4692–4708 (2010)

Kellerer, A.M.: Considerations on the random traversal of convex bodies and solutions for general cylinders. Radiat. Res. **47**(2), 359–376 (1971)

Kellerer, A.M., Rossi, H.H.: The theory of dual radiation action. Curr. Top. Radiat. Res. Q. **8**, 85–158 (1972)

Kellerer, A.M., Chmelevsky D.: Concepts of microdosimetry. II. Probability distributions of the microdosimetric variables. Radiat. Environ. Biophys. **12**(3), 205–216 (1975)

Lappa, A.V., Bigildeev, E.A., Burmistrov, D.S., Vassiliev, O.N.: Trion code for radiation action calculations and its application in microdosimetry and radiobiology. Radiat. Environ. Biophys. **32**(1), 1–19 (1993)

Nikjoo, H., Goodhead D.T., Charlton, D.E., Paretzke, H.G.: Energy deposition in small cylindrical targets by ultrasoft X-rays. Phys. Med. Biol. **34**(6), 691–705 (1989)

Nikjoo, H., Goodhead D.T., Charlton, D.E., Paretzke, H.G.: Energy deposition in small cylindrical targets by monoenergetic electrons. Int. J. Radiat. Biol. **60**(5), 739–756 (1991)

Nikjoo, H., O'Neill, P., Goodhead, D.T., Terrisol, M.: Computational modelling of low-energy electron-induced DNA damage by early physical and chemical events. Int. J. Radiat. Biol. **71**(5), 467–483 (1997)

Nikjoo, H., Uehara, S., Emfietzoglou, D, Cucinotta, F.A.: Track-structure codes in radiation research. Radiat. Meas. **41**(9–10), 1052–1074 (2006)

Palajova, Z., Spurny, F., Davidkova, M.: Microdosimetry distributions for 40–200 MeV protons. Radiat. Protect. Dosim. **121**(4), 376–381 (2006)

Rossi, H.H., Rosenzweig, W.: Measurements of neutron dose as a function of linear energy transfer. Radiat. Res. **2**(5), 417–425 (1955)

Rossi, H.H., Zaider, M.: Microdosimetry and Its Applications. Springer, Berlin (1996)

Saito, K., Otsuka, M.: Theory of statistical fluctuations in neutron distributions. J. Nucl. Sci. Technol. **2**(8), 304–314 (1965)

Turner, J.E., Paretzke, H.G., Hamm, H.G., Wright, H.A., Ritchie, R.H.: Comparative study of electron energy deposition and yields in water in the liquid and vapor phases. Radiat. Res. **92**, 47–60 (1982)

Vassiliev, O.N.: The influence of short tracks on the biological effectiveness of radiations. Radiat. Protect. Dosim. **62**(3), 179–182 (1995)

Wang, H., Vassiliev O.N.: Microdosimetric characterisation of radiation fields for modelling tissue response in radiotherapy. Int. J. Cancer Ther. Oncol. **2**(1), 020116 (2014)

Zaider, M., Brenner D.J.: On the microdosimetric definition of quality factors. Radiat. Res. **103**(3), 302–316 (1985)

Chapter 7
Grid Based Boltzmann Equation Solvers

7.1 Beyond the Monte Carlo method

The Monte Carlo method is not the only method for solving the Boltzmann transport equation. Furthermore, the popular belief that Monte Carlo is the most accurate method for solving problems of radiation transport is simply not correct. Non-stochastic or deterministic methods for solving the Boltzmann equation achieve the same level of accuracy as Monte Carlo algorithms. After all, both the Monte Carlo method and deterministic methods solve the same equation. In both methods a numerical solution converges to the exact solution. In Monte Carlo algorithms the convergence is stochastic. Increasing the number of histories brings a numerical Monte Carlo solution closer to the exact solution only in *probability* (Dudewicz and Mishra 1988, Appendix A). Monte Carlo algorithms may also have systematic errors associated with condensed history models, tracking cutoffs, etc. These can be reduced or eliminated, to achieve the desired overall accuracy. In deterministic algorithms, bringing a numerical solution closer to the exact solution involves increasing the resolution of computing grids and increasing the number of iterations in iterative procedures.

Deterministic algorithms were developed primarily for solving problems of neutron transport (Lewis and Miller 1993). These methods are often referred to as discrete ordinates methods, after a method for discretizing angular variables common to algorithms of this type. However, this terminology is somewhat misleading. You will see in this chapter that angular discretization is just one of several techniques used in a deterministic algorithm to find the solution. Furthermore, discrete ordinates is not the only method for discretizing the angular variables. For example, St. Aubin et al. (2016), developed an algorithm in which the discrete ordinates method was not used at all, instead the finite elements method was applied to both spatial and angular variables. For these reasons, when deterministic algorithms for solving the Boltzmann equation were introduced for dose calculations in radiotherapy, a more general terminology was adopted. Acuros® was the first

O.N. Vassiliev, *Monte Carlo Methods for Radiation Transport*, Biological and Medical Physics, Biomedical Engineering, DOI 10.1007/978-3-319-44141-2_7

algorithm of this type to be implemented in a commercial treatment planning system. When it was first introduced (Vassiliev et al. 2010), it was described as a grid based Boltzmann equation solver. Following the terminology of that paper, we will use the abbreviation GBBS for this type of deterministic solver.

Early experience with Acuros® in a clinical setting has been consistently positive, see for example, Bush et al. (2011), Hoffmann et al. (2012), Kan et al. (2012, 2013), Kroon et al. (2013), and Tsuruta et al. (2014). These results suggest that this deterministic approach has the potential to advance dose calculation algorithms for treatment planning in radiotherapy. Further evidence of that potential is a new algorithm for solving the Boltzmann equation in the presence of external magnetic fields developed by St. Aubin et al. (2015) and (2016). This algorithm is applicable to dose calculations for MRI-guided radiotherapy systems.

This justifies, we think, including this topic in a Monte Carlo book. It will give readers a better perspective. We will present the material in a rather simplified and concise manner. For a more comprehensive introduction to deterministic methods, we refer to the classic book by Lewis and Miller (1993).

For simplicity we will assume that there is only one type of particle, and no secondary particles are produced. We will not consider any condensed history models, even though these may be necessary for electron transport in radiotherapy applications. Finally, as in the Monte Carlo method, multiple options are available in the GBBS method for designing an optimal algorithm. In this chapter, we will give precedence to simpler options.

7.2 Discretize, Discretize, Discretize

Discretization of the phase space is an important step in designing a GBBS algorithm. In this section we will consider discretization methods separately for the energy, angular, and spatial variables.

7.2.1 Discretization of Energy: The Multigroup Approximation

The method for discretizing the particle energy is called multigroup approximation. First, we need to determine the range of energies in which we are interested. Often it is the interval from the maximum energy of the source to the tracking cut off energy. Then, we divide this energy range into smaller intervals, with the interval boundaries $\{E_0, E_1, \ldots, E_G\}$, and interval widths $\Delta E_g = E_{g-1} - E_g$, $g = 1, 2, \ldots, G$. The common convention is that E_0 is the maximum energy, $E_{g-1} > E_g$, and E_G is the minimum energy. The energy interval from E_{g-1} to E_g is referred to as energy group g. The parameter G defines the total number of energy groups. Choosing a larger G improves the accuracy of the algorithm, but increases the computing time.

Next, we choose the form of the Boltzmann equation corresponding to the physical problem that we need to solve. We choose the simple form given by Eq. (3.40):

$$\left(\vec{\Omega}\cdot\vec{\nabla}\right)\Phi\left(\vec{r},\vec{\Omega},E\right)+\sigma\left(\vec{r},E\right)\Phi\left(\vec{r},\vec{\Omega},E\right)$$
$$=S\left(\vec{r},\vec{\Omega},E\right)+\int \mathrm{d}E'\int \mathrm{d}\vec{\Omega}'\sigma_s\left(\vec{r},\vec{\Omega}',E'\rightarrow\vec{\Omega},E\right)\Phi\left(\vec{r},\vec{\Omega}',E'\right). \quad (7.1)$$

Then, in the collision integral we replace the integral over E' with the sum of integrals over all energy groups:

$$\int \mathrm{d}E'\ldots=\sum_{g'=1}^{G}\int_{E_{g'}}^{E_{g'-1}}\mathrm{d}E'\ldots \quad (7.2)$$

Next, we integrate Eq. (7.1) over energy group g, that is from E_g to E_{g-1}. The integration is straightforward. It brings us to the following form of the Boltzmann equation in the multigroup approximation:

$$\left(\vec{\Omega}\cdot\vec{\nabla}\right)\Phi_g\left(\vec{r},\vec{\Omega}\right)+\sigma_g\left(\vec{r}\right)\Phi_g\left(\vec{r},\vec{\Omega}\right)$$
$$=S_g\left(\vec{r},\vec{\Omega}\right)+\sum_{g'=1}^{G}\int \mathrm{d}\vec{\Omega}'\Phi_{g'}\left(\vec{r},\vec{\Omega}'\right)\sigma_s\left(\vec{r},\vec{\Omega}\cdot\vec{\Omega}',g'\rightarrow g\right), \quad (7.3)$$

where we introduced the notations

$$\Phi_g\left(\vec{r},\vec{\Omega}\right)=\int_{E_g}^{E_{g-1}}\mathrm{d}E\ \Phi\left(\vec{r},\vec{\Omega},E\right). \quad (7.4)$$

$$S_g\left(\vec{r},\vec{\Omega}\right)=\int_{E_g}^{E_{g-1}}\mathrm{d}E\ S\left(\vec{r},\vec{\Omega},E\right). \quad (7.5)$$

$$\sigma_g\left(\vec{r}\right)=\frac{1}{\Phi_g\left(\vec{r},\vec{\Omega}\right)}\int_{E_g}^{E_{g-1}}\mathrm{d}E\ \sigma\left(\vec{r},E\right)\Phi\left(\vec{r},\vec{\Omega},E\right). \quad (7.6)$$

$$\sigma_s\left(\vec{r},\vec{\Omega}\cdot\vec{\Omega}',g'\rightarrow g\right)$$
$$=\frac{1}{\Phi_{g'}\left(\vec{r},\vec{\Omega}'\right)}\int_{E_g}^{E_{g-1}}\mathrm{d}E\int_{E_{g'}}^{E_{g'-1}}\mathrm{d}E'\sigma_s\left(\vec{r},\vec{\Omega}',E'\rightarrow\vec{\Omega},E\right)\Phi\left(\vec{r},\vec{\Omega}',E'\right). \quad (7.7)$$

Implicit in these definitions is the assumption that the ratio $\Phi(\vec{r}, \vec{\Omega}, E) \,/\, \Phi_g(\vec{r}, \vec{\Omega})$ is independent of $\vec{r}$ and $\vec{\Omega}$. This assumption is known as the energy separability approximation and can be expressed as follows

$$\Phi\left(\vec{r}, \vec{\Omega}, E\right) \approx f\,(E)\,\tilde{\Phi}_g\left(\vec{r}, \vec{\Omega}\right);\;\; E_g < E \leq E_{g-1}. \tag{7.8}$$

Another important observation is that the coefficients $\sigma_g(\vec{r})$ and $\sigma_s(\vec{r}, \vec{\Omega}\cdot\vec{\Omega}\,', g' \rightarrow g)$, according to their definitions, are functionals of the fluence spectrum, $f(E)$. The latter, of course, is not known until the Boltzmann equation is solved. This difficulty is resolved by using an approximate fluence spectrum in Eqs. (7.6) and (7.7). For example, the fluence spectrum can be approximated by a piecewise constant function:

$$f\,(E) \approx \frac{1}{\Delta E_g},\;\; E_g < E \leq E_{g-1}. \tag{7.9}$$

Uncertainties associated with this approximation can be reduced to an acceptable level by reducing the widths ΔE_g of the energy groups. The group cross section $\sigma_g(\vec{r}\,)$ can be pre-tabulated, for each material in the problem. The scattering cross section $\sigma_s(\vec{\Omega}\cdot\vec{\Omega}\,', g' \rightarrow g)$ is more difficult to tabulate because of its angular dependence. This will become an easier task after we use a spherical harmonics expansion to simplify the collision integral.

7.2.2 *Spherical Harmonics Series for the Collision Integral*

Before discretizing the angular variables, we will transform the collision integral into a more convenient form. First, we approximate the angular dependence of the fluence by a spherical harmonics expansion:

$$\Phi_g\left(\vec{r}, \vec{\Omega}\right) \approx \sum_{l=0}^{L}\sum_{m=-l}^{l} \Phi_{gl}^{m}\left(\vec{r}\,\right) Y_l^m\left(\vec{\Omega}\right), \tag{7.10}$$

where $Y_l^m(\vec{\Omega})$ is a spherical harmonic and L is an algorithm parameter. Increasing L improves the accuracy, at the expense of an increased computing time. The spherical harmonics are orthogonal (Arfken et al. 2013):

$$\int \mathrm{d}\vec{\Omega}\; Y_l^{m*}\left(\vec{\Omega}\right) Y_{l'}^{m'}\left(\vec{\Omega}\right) = \delta_{ll'}\delta_{mm'}, \tag{7.11}$$

where '$*$' denotes the complex conjugate. From Eq. (7.11) we obtain the expression for the coefficients of the expansion

$$\Phi_{gl}^{m}\left(\vec{r}\right) = \int \mathrm{d}\vec{\Omega}\; Y_{l}^{m*}\left(\vec{\Omega}\right) \Phi_{g}\left(\vec{r},\vec{\Omega}\right). \tag{7.12}$$

The fluence is generally a function of two angular variables, because a direction in three dimensions is defined by two angles, such as the polar angle θ and the azimuthal angle ϕ in the spherical coordinate system. In contrast, the scattering cross section usually depends on only one angle, the scattering angle, or, in our notation its cosine, $\vec{\Omega}\cdot\vec{\Omega}\,'$. This implies azimuthally symmetric scattering, an approximation that is unlikely to be challenged in the context of radiotherapy applications. Hence, for the angular dependence of the scattering cross section, instead of spherical harmonics, we can use the simpler Legendre series

$$\sigma_{s}\left(\vec{r},\vec{\Omega}\cdot\vec{\Omega}\,',g'\rightarrow g\right) \approx \sum_{l=0}^{L} \sigma_{sl}\left(\vec{r},g'\rightarrow g\right) P_{l}\left(\vec{\Omega}\cdot\vec{\Omega}\,'\right), \tag{7.13}$$

where $P_l(\vec{\Omega}\cdot\vec{\Omega}\,')$ is a Legendre polynomial. Legendre polynomials are also orthogonal (Arfken et al. 2013):

$$\int_{-1}^{1} \mathrm{d}\mu\; P_{l}\left(\mu\right) P_{l'}\left(\mu\right) = \frac{2}{2l+1}\delta_{ll'}, \tag{7.14}$$

and the formula for the coefficients of the expansion that follows from the orthogonality condition is:

$$\sigma_{sl}\left(\vec{r},g'\rightarrow g\right) = \frac{2l+1}{2}\int_{-1}^{1} \mathrm{d}\mu\; P_{l}\left(\mu\right)\sigma_{s}\left(\vec{r},\mu,g'\rightarrow g\right). \tag{7.15}$$

Next, we insert Eqs. (7.10) and (7.13) into the collision integral in Eq. (7.3). The collision integral becomes

$$\sum_{g'=1}^{G}\sum_{l=0}^{L}\sum_{m=-l}^{l}\sum_{l'=0}^{L} \Phi_{g'l}^{m}\left(\vec{r}\right)\sigma_{sl'}\left(\vec{r},g'\rightarrow g\right)\int \mathrm{d}\vec{\Omega}\,'\; Y_{l}^{m}\left(\vec{\Omega}\,'\right) P_{l'}\left(\vec{\Omega}\cdot\vec{\Omega}\,'\right). \tag{7.16}$$

Then, we use the addition theorem (Arfken et al. 2013)

$$P_{l'}\left(\vec{\Omega}\cdot\vec{\Omega}\,'\right) = \frac{4\pi}{2l'+1}\sum_{m'=-l'}^{l'} Y_{l'}^{m'}\left(\vec{\Omega}\right) Y_{l'}^{m'*}\left(\vec{\Omega}\,'\right) \tag{7.17}$$

to replace the Legendre polynomial with spherical harmonics and simplify the integral:

$$\int \mathrm{d}\vec{\Omega}'\, Y_l^m\left(\vec{\Omega}'\right) P_{l'}\left(\vec{\Omega}\cdot\vec{\Omega}'\right)$$

$$= \frac{4\pi}{2l'+1}\sum_{m'=-l'}^{l'}\int \mathrm{d}\vec{\Omega}'\, Y_l^m\left(\vec{\Omega}'\right) Y_{l'}^{m'}\left(\vec{\Omega}\right) Y_{l'}^{m'*}\left(\vec{\Omega}'\right) = \frac{4\pi}{2l'+1} Y_{l'}^m\left(\vec{\Omega}\right)\delta_{ll'}. \tag{7.18}$$

At the last step in Eq. (7.18), in the integral over $\vec{\Omega}'$, we used the orthogonality of spherical harmonics, Eq. (7.11). Finally, as we insert this result into Eq. (7.16), the factor $\delta_{ll'}$ eliminates the sum over l'. This brings us to a form of the Boltzmann equation in which the collision integral is approximated by a finite sum:

$$\left(\vec{\Omega}\cdot\vec{\nabla}\right)\Phi_g\left(\vec{r},\vec{\Omega}\right) + \sigma_g\left(\vec{r}\right)\Phi_g\left(\vec{r},\vec{\Omega}\right)$$

$$= S_g\left(\vec{r},\vec{\Omega}\right) + \sum_{l=0}^{L}\frac{4\pi}{2l+1}\sum_{g'=1}^{G}\sigma_{sl}\left(\vec{r},g'\to g\right)\sum_{m=-l}^{l}\Phi_{g'l}^m\left(\vec{r}\right) Y_l^m\left(\vec{\Omega}\right). \tag{7.19}$$

The coefficients $\sigma_{sl}\left(\vec{r},g'\to g\right)$ defining the doubly differential scattering cross sections can be tabulated. The table is three dimensional, and the indices are: l, g', and g.

7.2.3 *Discretization of Angular Variables: The Discrete Ordinates Method*

The most common technique for discretizing the angular variables is the discrete ordinates method. It is also the simplest. In this technique the solution, $\Phi_g(\vec{r},\vec{\Omega})$, is calculated only for a finite discrete set of directions $\vec{\Omega}_{ij} = (\mu_i,\phi_j)$, $i = 1,2,\ldots,n_\mu$, $j = 1,2,\ldots,n_\phi$. The discrete directions $\vec{\Omega}_{ij}$ are chosen to optimize numerical integration of two-dimensional integrals over $\vec{\Omega}$, such as the integral in Eq. (7.12). As an example of a quadrature that works well in the context of the Boltzmann equation, we will describe the Chebyshev-Legendre quadrature. The use of this quadrature in discrete ordinate algorithms is discussed by Walters (1987). The idea of the integration method is that in a two-dimensional integral, integration over one variable is performed using the Gauss-Chebyshev quadrature and integration over the other variable is performed using the Gauss-Legendre quadrature. Both quadratures are covered in most textbooks on numerical methods. The numerical data needed for the implementation of both quadratures, as well as the integration errors, can be found, for example, in Abramowitz and Stegan (1964).

Problem

Calculate the following integral:

$$I = \int \mathrm{d}\vec{\Omega}\, f\left(\vec{\Omega}\right). \tag{7.20}$$

Solution

We express the integral in terms of the angular variables $\mu = \cos\theta$ and ϕ:

$$I = \int_{-1}^{1} \mathrm{d}\mu \int_{0}^{2\pi} \mathrm{d}\phi\, f(\mu, \phi). \tag{7.21}$$

We break the second integral into two integrals, from 0 to π and from π to 2π and use the substitution $\nu = \cos\phi$. Consequently, we will have $\mathrm{d}\phi = -\mathrm{d}\nu/\sqrt{1-\nu^2}$ in the first integral ($0 \leq \phi < \pi$), and $\mathrm{d}\phi = \mathrm{d}\nu/\sqrt{1-\nu^2}$, in the second integral($\pi \leq \phi < 2\pi$). This produces two identical integrals, and the result is

$$I = 2\int_{-1}^{1} \mathrm{d}\mu \int_{-1}^{1} \mathrm{d}\nu\, \frac{f(\mu, \nu)}{\sqrt{1-\nu^2}}. \tag{7.22}$$

In the case of one-dimensional integrals, the Gauss-Legendre quadrature is applied as follows:

$$\int_{-1}^{1} \mathrm{d}\mu\, f(\mu) \approx \sum_{i=1}^{n_\mu} w_i^{(L)} f(\mu_i). \tag{7.23}$$

The abscissas μ_i are the zeros of Legendre polynomials. Tables of the abscissas and the weights $w_i^{(L)}$ are available in the literature, for example in Abramowitz and Stegan (1964). Similarly, the Gauss-Chebyshev quadrature for one-dimensional integrals is given by

$$\int_{-1}^{1} \mathrm{d}\nu\, \frac{f(\nu)}{\sqrt{1-\nu^2}} \approx \sum_{j=1}^{n_\phi} w_j^{(C)} f(\nu_j). \tag{7.24}$$

In this case

$$\nu_j = \cos\frac{(2j-1)\,\pi}{2n_\phi}, \tag{7.25}$$

and

$$w_j^{(C)} = \frac{\pi}{n_\phi}. \tag{7.26}$$

The Gauss-Chebyshev quadrature should not be confused with Chebyshev's equal weight integration formula.

Finally, combination of the two one-dimensional quadratures in the two-dimensional integral produce the following result

$$I \approx \sum_{i=1}^{n_\mu} \sum_{j=1}^{n_\phi} w_i^{(L)} w_j^{(C)} f\left(\vec{\Omega}_{ij}\right), \tag{7.27}$$

where $\vec{\Omega}_{ij} \equiv (\mu_i, \nu_j)$. If $n_\mu = n_\phi$, the quadrature is called the square Legendre-Chebyshev quadrature.

In the discrete ordinates method, the Boltzmann equation is simply written and solved for each of the discrete directions, $\vec{\Omega}_{ij}$:

$$\begin{aligned}
&\left(\vec{\Omega}_{ij} \cdot \vec{\nabla}\right) \Phi_g\left(\vec{r}, \vec{\Omega}_{ij}\right) + \sigma_g\left(\vec{r}\,\right) \Phi_g\left(\vec{r}, \vec{\Omega}_{ij}\right) \\
&\quad = S_g\left(\vec{r}, \vec{\Omega}_{ij}\right) + \sum_{l=0}^{L} \frac{4\pi}{2l+1} \sum_{g'=1}^{G} \sigma_{sl}\left(\vec{r}, g' \to g\right) \sum_{m=-l}^{l} \Phi_{g'l}^m\left(\vec{r}\,\right) Y_l^m\left(\vec{\Omega}_{ij}\right), \\
&\quad i = 1, 2, \ldots, n_\mu; \;\; j = 1, 2, \ldots, n_\phi, \;\; g = 1, 2, \ldots, G,
\end{aligned} \tag{7.28}$$

with a two-dimensional quadrature, such as the Chebyshev-Legendre quadrature, which is used to calculate the coefficients $\Phi_{g'l}^m$:

$$\begin{aligned}
\Phi_{g'l}^m\left(\vec{r}\,\right) &= \int \mathrm{d}\vec{\Omega} \;\; Y_l^{m*}\left(\vec{\Omega}\right) \Phi_{g'}\left(\vec{r}, \vec{\Omega}\right) \\
&\approx \sum_{i'=1}^{n_\mu} \sum_{j'=1}^{n_\phi} w_{i'}^{(L)} w_{j'}^{(C)} Y_l^{m*}\left(\vec{\Omega}_{i'j'}\right) \Phi_{g'}\left(\vec{r}, \vec{\Omega}_{i'j'}\right).
\end{aligned} \tag{7.29}$$

For each energy group g, we have a system of $n_\mu \times n_\phi$ equations. The equations are coupled, because the equation for any discrete direction $\vec{\Omega}_{ij}$ contains fluences for all the other directions, in the sum in Eq. (7.29). These equations are differential, because of the operator ∇ in the streaming operator. We will now discuss one of the standard numerical methods for solving differential equations, the finite elements method. Applying this method to Eq. (7.28) will transform the system of partial differential equations, into an even larger system of linear algebraic equations.

7.2.4 *Discretization of Spatial Variables: The Finite Elements Method*

The first step is defining the computational grid for the spatial variables, x, y and z. This process is similar to that of voxelizing a phantom for Monte Carlo calculations. The resolution of the grid should be higher where large gradients of the solution are expected, near the beam edges, in high density gradient regions, and so on. Here we will consider a tetrahedral grid. Each tetrahedron of the grid represents a finite element. As an example, a part of a tetrahedral mesh superimposed on the head anatomy is shown in Fig.7.1.

Let us consider one tetrahedron. It has four vertices located at points $\vec{r}_{k'}$, $k' = 1, 2, 3, 4$. They are also referred to as the nodes. With each of the four vertices we will associate a basis function $h_k(x, y, z)$, $k = 1, 2, 3, 4$. The basis functions satisfy the following conditions:

$$h_k\left(\vec{r}_{k'}\right) = \delta_{kk'};\ \ k = 1, 2, 3, 4;\ \ k' = 1, 2, 3, 4. \tag{7.30}$$

It means, for example, that the basis function h_3 has the value of 1 on node 3 and is zero on all the other nodes. We will use linear basis functions

$$h_k\left(x, y, z\right) = a_k x + b_k y + c_k z + d_k;\ \ k = 1, 2, 3, 4, \tag{7.31}$$

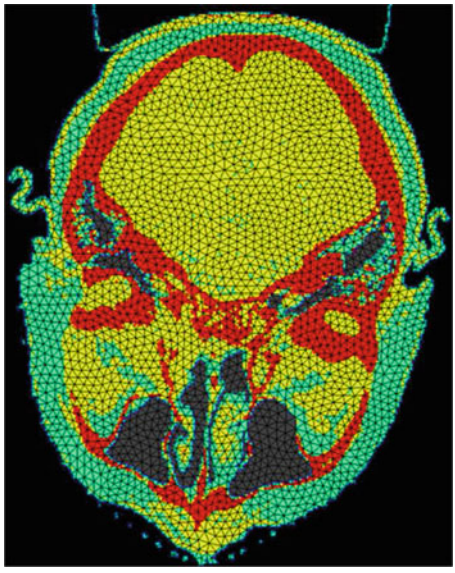

Fig. 7.1 Tetrahedral grid superimposed on head anatomy, (Vassiliev et al. 2008). Used with permission

where the coefficients, a_k, b_k, c_k, and d_k, can be found by solving the system of equation given by Eq. (7.30). Within the tetrahedron, we will seek a solution of the Boltzmann equation, Eq. (7.28), in the following form:

$$\Phi_g\left(\vec{r},\vec{\Omega}_{ij}\right)=\sum_{k=1}^{4}h_k\left(\vec{r}\right)\Phi_{gk}\left(\vec{\Omega}_{ij}\right). \tag{7.32}$$

It follows from Eq. (7.30) that the coefficients $\Phi_{gk}(\vec{\Omega}_{ij})$ are nothing but the fluence at the nodes of the grid, that is, at the vertices of the tetrahedron: $\Phi_{gk}(\vec{\Omega}_{ij}) = \Phi_g(\vec{r}_k,\vec{\Omega}_{ij})$. Then, Eq. (7.32) is simply a formula for linear interpolation in three dimensions. Our next task is to derive an equation for the unknown coefficients $\Phi_{gk}(\vec{\Omega}_{ij})$. For that purpose we will use the Galerkin method.

7.3 The Galerkin Method

First, we will present the Galerkin method in a general form. Then, we will apply the method to the discretized Boltzmann equation, Eq. (7.28).

Let us consider the following equation for an unknown function $\Phi(x)$:

$$\hat{L}\Phi\left(x\right)=S\left(x\right), \tag{7.33}$$

where $\hat{L}$ is an arbitrary linear operator, the source function $S(x)$ is given, and x is a multidimensional variable. We need to find a solution $\Phi(x)$. If our numerical solution is $\tilde{\Phi}(x)$, then the residual $R(x)$ is defined as follows:

$$R\left(x\right)=\hat{L}\tilde{\Phi}\left(x\right)-S\left(x\right). \tag{7.34}$$

If we were able to solve the problem exactly, then the residual would be equal to zero for all x. We, however, will set a more modest goal of finding an approximate solution $\tilde{\Phi}(x)$ that in "some sense" minimizes the residual. In the Galerkin method, to achieve this goal, the following N conditions have to be satisfied:

$$\int R\left(x\right)h_k\left(x\right)\mathrm{d}x=0;\;\; k=1,2,\ldots,N. \tag{7.35}$$

The functions $h_k(x)$ are referred to as the weighting functions. We, however, will use here the basis functions that we introduced in the preceding section. For this reason we used notations h_k for the weighting functions. These N conditions are described in the literature as the orthogonality of the residual to a set of weighting functions, $h_1,\ldots,h_N$.

We need to clarify in what sense satisfying the orthogonality condition minimizes the residual. To do so, we will approximate the residual by a linear combination of basis functions:

$$R(x) \approx \sum_{k=1}^{N} a_k h_k(x), \tag{7.36}$$

where a_k are unknown coefficients. To find the coefficients, we will perform a least squares fit by minimizing the following integral:

$$\int \left[R(x) - \sum_{k=1}^{N} a_k h_k(x) \right]^2 \mathrm{d}x. \tag{7.37}$$

To find the minimum, we take the derivatives $\partial/\partial a_{k'}$ ($k' = 1, 2, \ldots, N$) of this integral and require that they are all equal to zero. This produces the following system of N linear algebraic equations for coefficients a_k :

$$\int R(x) h_{k'}(x) \, \mathrm{d}x = \sum_{k=1}^{N} a_k \int h_{k'}(x) h_k(x) \, \mathrm{d}x. \tag{7.38}$$

If the residual $R(x)$ is orthogonal to the set of basis functions h_k, then the left-hand side of the equation is equal to zero for all k'. That, in turn, means that all coefficients a_k are also equal to zero, provided that the matrix

$$\int h_{k'}(x) h_k(x) \, \mathrm{d}x. \tag{7.39}$$

is not singular, in other words, that its determinant is not zero. With all coefficients a_k equal to zero, the approximate expression for the residual given by Eq. (7.36) is also equal to zero for all x. This means that, if the orthogonality conditions Eq. (7.35) are satisfied, then the least squares fit to the residual is zero for all x.

Before applying the Galerkin method to the Boltzmann equation, we recall that for each tetrahedron, each discrete direction $\vec{\Omega}_{ij}$, and each energy group g we need to find four coefficients $\Phi_{gk}(\vec{\Omega}_{ij})$, $k = 1, 2, 3, 4$. These coefficients are the fluences at the four vertices of the tetrahedron. This means that we need to transform Eq. (7.28) for each set of indices (i, j, g) into a system of four equations. We achieve this by multiplying Eq. (7.28) by basis functions h_1, h_2, h_3, and h_4, one at a time, and then integrating the equation over the volume of the tetrahedron. This procedure is equivalent to satisfying the orthogonality condition for the residual. That is, $R(x)$ is orthogonal to the set of four functions: h_1, h_2, h_3, h_4. Integration will eliminate the derivatives in the streaming operator, after we apply the divergence theorem to this term, which ultimately will yield a large system of linear algebraic equations. We will demonstrate this technique for one tetrahedron and one basis function, h_1.

We multiply Eq. (7.28) by $h_1(x)$ and integrate the result over the volume V of the tetrahedron. We will integrate one term of the equation at a time, starting with the simplest term, the removal operator:

$$\int_V \mathrm{d}\vec{r}\, h_1\left(\vec{r}\right) \sigma_g\left(\vec{r}\right) \Phi_g\left(\vec{r}, \vec{\Omega}_{ij}\right) = \sigma_g \int_V \mathrm{d}\vec{r}\, h_1\left(\vec{r}\right) \sum_{k=1}^{4} h_k\left(\vec{r}\right) \Phi_{gk}\left(\vec{\Omega}_{ij}\right)$$

$$= \sigma_g \sum_{k=1}^{4} \Phi_{gk}\left(\vec{\Omega}_{ij}\right) \int_V \mathrm{d}\vec{r}\, h_1\left(\vec{r}\right) h_k\left(\vec{r}\right) \equiv \sigma_g \sum_{k=1}^{4} \Phi_{gk}\left(\vec{\Omega}_{ij}\right) H_{1k}^{(V)}. \quad (7.40)$$

We assumed above that the material within the finite element is homogeneous. In that case the cross section is constant within the integration volume, V. Then, we inserted Eq. (7.32) for Φ_g, and introduced the notation $H_{1k}^{(V)}$ for the integral of the product of two basis functions.

Integration of the source term is very similar, so we present only the result:

$$\int_V \mathrm{d}\vec{r}\, h_1\left(\vec{r}\right) S_g\left(\vec{r}, \vec{\Omega}_{ij}\right) = \sum_{k=1}^{4} S_{gk}\left(\vec{\Omega}_{ij}\right) H_{1k}^{(V)}. \quad (7.41)$$

We need to mention only that like the fluence, the source function was approximated by a sum, as in Eq. (7.32).

Similarly, we integrate the collision integral:

$$\int_V \mathrm{d}\vec{r}\, h_1\left(\vec{r}\right) \sum_{l=0}^{L} \frac{4\pi}{2l+1} \sum_{g'=1}^{G} \sigma_{sl}\left(g' \to g\right) \sum_{m=-l}^{l} \Phi_{g'l}^{m}\left(\vec{r}\right) Y_l^m\left(\vec{\Omega}_{ij}\right)$$

$$= \sum_{l=0}^{L} \frac{4\pi}{2l+1} \sum_{g'=1}^{G} \sigma_{sl}\left(g' \to g\right) \sum_{m=-l}^{l} Y_l^m\left(\vec{\Omega}_{ij}\right) \sum_{k=1}^{4} \Phi_{g'lk}^{m} H_{1k}^{(V)}. \quad (7.42)$$

Finally, we integrate the most difficult part, the streaming operator:

$$\int_V \mathrm{d}\vec{r}\, h_1\left(\vec{r}\right) \left(\vec{\Omega}_{ij} \cdot \vec{\nabla}\right) \Phi_g\left(\vec{r}, \vec{\Omega}_{ij}\right)$$

$$= \int_V \mathrm{d}\vec{r}\, \left(\vec{\Omega}_{ij} \cdot \vec{\nabla}\right) \left[h_1\left(\vec{r}\right) \Phi_g\left(\vec{r}, \vec{\Omega}_{ij}\right)\right] - \int_V \mathrm{d}\vec{r}\, \Phi_g\left(\vec{r}, \vec{\Omega}_{ij}\right) \left(\vec{\Omega}_{ij} \cdot \vec{\nabla}\right) h_1\left(\vec{r}\right). \quad (7.43)$$

In the first integral on the right-hand side, we apply the divergence theorem

$$\int_V \mathrm{d}\vec{r}\, \left(\vec{\Omega}_{ij} \cdot \vec{\nabla}\right) \left[h_1\left(\vec{r}\right) \Phi_g\left(\vec{r}, \vec{\Omega}_{ij}\right)\right] = \oint_{\Gamma(V)} \mathrm{d}\Gamma \left(\vec{\Omega}_{ij} \cdot \vec{n}_\Gamma\right) h_1\left(\vec{r}\right) \Phi_g\left(\vec{r}, \vec{\Omega}_{ij}\right), \quad (7.44)$$

where $\Gamma(V)$ is the total surface area of the tetrahedron, that is, the area of all four faces, and $\vec{n}_\Gamma = \vec{n}_\Gamma(\vec{r})$ is the outward pointing normal to Γ, at point $\vec{r}$ on the surface. The integral over the surface of the tetrahedron, $\Gamma(V)$, is equal to the sum of integrals over the four faces, Γ_s, $s = 1, 2, 3, 4$:

$$\oint_{\Gamma(V)} \mathrm{d}\Gamma \left(\vec{\Omega}_{ij} \cdot \vec{n}_{\Gamma}\right) h_1\left(\vec{r}\,\right) \Phi_g\left(\vec{r}, \vec{\Omega}_{ij}\right)$$
$$= \sum_{s=1}^{4} \left(\vec{\Omega}_{ij} \cdot \vec{n}_{\Gamma s}\right) \int_{\Gamma_s} \mathrm{d}\Gamma\, h_1\left(\vec{r}\,\right) \Phi_g\left(\vec{r}, \vec{\Omega}_{ij}\right). \tag{7.45}$$

Finally, we yet again approximate the fluence by a sum, as in Eq. (7.32):

$$\sum_{s=1}^{4} \left(\vec{\Omega}_{ij} \cdot \vec{n}_{\Gamma s}\right) \int_{\Gamma_s} \mathrm{d}\Gamma\, h_1\left(\vec{r}\,\right) \Phi_g\left(\vec{r}, \vec{\Omega}_{ij}\right)$$
$$= \sum_{k=1}^{4} \Phi_{gk}\left(\vec{\Omega}_{ij}\right) \sum_{s=1}^{4} \left(\vec{\Omega}_{ij} \cdot \vec{n}_{\Gamma s}\right) \int_{\Gamma_s} \mathrm{d}\Gamma\;\; h_1\left(\vec{r}\,\right) h_k\left(\vec{r}\,\right)$$
$$\equiv \sum_{k=1}^{4} \Phi_{gk}\left(\vec{\Omega}_{ij}\right) \sum_{s=1}^{4} \left(\vec{\Omega}_{ij} \cdot \vec{n}_{\Gamma s}\right) H_{1k}^{(s)}, \tag{7.46}$$

where we introduced the notation $H_{1k}^{(s)}$ for the integral over face s of the product of two basis functions. Integration over tetrahedron faces should be performed carefully. If continuity of the fluence across the surface of a finite element is not required, which is the case in discontinuous finite element algorithms, the fluence may have two different values on the two sides of each face of the tetrahedron. For example, if tetrahedron 1 shares face s with tetrahedron 2, then we can identify one side of surface s as side s_1 that belongs to tetrahedron 1, and the other, as side s_2. Then, when writing an equation for tetrahedron 1, the conventional approach is that for all directions $\vec{\Omega}_{ij}$ pointing into tetrahedron 1 ($\vec{\Omega}_{ij} \cdot \vec{n}_{\Gamma s} < 0$), for points on surface s, the fluence on side s_2 of the surface is used. In that case, the fluence on s_2 is called the incoming fluence. Conversely, if ($\vec{\Omega}_{ij} \cdot \vec{n}_{\Gamma s} > 0$), then the fluence on side s_1 is used.

We, however, have not yet finished with the streaming operator. There is a second integral in Eq. (7.43). In this case, the algebra is simple:

$$\int_V \mathrm{d}\vec{r}\, \Phi_g\left(\vec{r}, \vec{\Omega}_{ij}\right) \left(\vec{\Omega}_{ij} \cdot \vec{\nabla}\right) h_1\left(\vec{r}\,\right)$$
$$= \sum_{k=1}^{4} \Phi_{gk}\left(\vec{\Omega}_{ij}\right) \int_V \mathrm{d}\vec{r}\, h_k\left(\vec{r}\,\right) \left(\vec{\Omega}_{ij} \cdot \vec{\nabla}\right) h_1\left(\vec{r}\,\right) \equiv \sum_{k=1}^{4} \Phi_{gk}\left(\vec{\Omega}_{ij}\right) H_{1k}^{(ij)}, \tag{7.47}$$

where we introduced an obvious notation $H_{1k}^{(ij)}$ for the integral. Combining Eqs. (7.40)–(7.42), (7.46), and (7.47) we arrive at a discretized form of the Boltzmann equation, where the spatial derivatives have been eliminated by applying the Galerkin method:

$$\sum_{k=1}^{4}\Phi_{gk}\left(\vec{\Omega}_{ij}\right)\sum_{s=1}^{4}\left(\vec{\Omega}_{ij}\cdot\vec{n}_{\Gamma s}\right)H_{1k}^{(s)}-\sum_{k=1}^{4}\Phi_{gk}\left(\vec{\Omega}_{ij}\right)H_{1k}^{(ij)}$$
$$+\sigma_g\sum_{k=1}^{4}\Phi_{gk}\left(\vec{\Omega}_{ij}\right)H_{1k}^{(V)}=\sum_{k=1}^{4}S_{gk}\left(\vec{\Omega}_{ij}\right)H_{1k}^{(V)}$$
$$+\sum_{l=0}^{L}\frac{4\pi}{2l+1}\sum_{g'=1}^{G}\sigma_{sl}\left(g'\rightarrow g\right)\sum_{m=-l}^{l}Y_l^m\left(\vec{\Omega}_{ij}\right)\sum_{k=1}^{4}\Phi_{g'lk}^{m}H_{1k}^{(V)}. \qquad (7.48)$$

To obtain four equations needed to find the fluences $\Phi_{gk}(\vec{\Omega}_{ij})$ at all four vertices of the tetrahedron, we have to repeat this exercise for the other three basis functions h_2, h_3, and h_4. The result, however, is predictable: we will get the same equation but with the subscript "1" replaced with 2, 3, and 4, respectively. To obtain a closed system of equations, we also need to express the coefficients $\Phi_{g'lk}^{m}$ on the right-hand side in terms of the fluences $\Phi_{gk}(\vec{\Omega}_{ij})$. To do so, we will use Eq. (7.29) written for a point $\vec{r}=\vec{r}_k$, keeping in mind that $\Phi_g(\vec{r}_k,\vec{\Omega}_{ij})=\Phi_{gk}(\vec{\Omega}_{ij})$:

$$\Phi_{g'lk}^{m}=\int \mathrm{d}\vec{\Omega}\;\; Y_l^{m*}\left(\vec{\Omega}\right)\Phi_{g'k}\left(\vec{\Omega}\right)\approx\sum_{i'=1}^{n_\mu}\sum_{j'=1}^{n_\phi}w_{i'}^{(L)}w_{j'}^{(C)}Y_l^{m*}\left(\vec{\Omega}_{i'j'}\right)\Phi_{g'k}\left(\vec{\Omega}_{i'j'}\right). \qquad (7.49)$$

In Eqs. (7.48) and (7.49) we now replace the subscript 1 with $k'=1,2,3,4$ and combine them in one equation that can be written in a matrix form:

$$\sum_{k=1}^{4}A_{gkk'}\left(\vec{\Omega}_{ij}\right)\Phi_{gk}\left(\vec{\Omega}_{ij}\right)$$
$$=S_{gk'}\left(\vec{\Omega}_{ij}\right)+\sum_{k=1}^{4}\sum_{g'=1}^{G}\sum_{i'=1}^{n_\mu}\sum_{j'=1}^{n_\phi}K_{kk'i'j'}\left(g'\rightarrow g,\vec{\Omega}_{i'j'}\cdot\vec{\Omega}_{ij}\right)\Phi_{g'k}\left(\vec{\Omega}_{i'j'}\right);$$
$$k'=1,2,3,4;\;\; g=1,2,\ldots,G;\;\; i=1,2,\ldots,n_\mu;\;\; j=1,2,\ldots,n_\phi, \qquad (7.50)$$

or in an operator form

$$\hat{A}\Phi=S+\hat{K}\Phi. \qquad (7.51)$$

Definitions of the matrices $A_{gkk'}(\vec{\Omega}_{ij})$, $S_{gk'}(\vec{\Omega}_{ij})$, and $K_{kk'i'j'}(g'\rightarrow g,\vec{\Omega}_{i'j'}\cdot\vec{\Omega}_{ij})$ are obvious from a comparison of Eq. (7.50) with Eqs. (7.48) and (7.49). The operators $\hat{A}$ and $\hat{K}$ are defined by the respective matrices, and Φ and S are matrices of size $G\times 4\times n_\mu\times n_\phi$ if only one tetrahedron is considered. This equation maintains the overall structure of the original, non-discretized Boltzmann equation.

The same procedure as above is followed when the equations are derived for a system comprised of many finite elements. It yields a large system of linear algebraic equations for the fluence coefficients $\Phi_{gk}\left(\vec{\Omega}_{ij}\right)$. In principle, the system can be solved by any of the standard numerical methods for this type of problems. The solution must satisfy boundary conditions. In the case of standard boundary conditions, Eq. (3.49), the incoming fluence is equal to zero on the tetrahedron faces forming the boundary of the computational domain. A very common approach to solving the system is described in the next section. It consists of a loop over energy groups (the outer loop) and an iterative procedure called source iterations (the inner loop).

7.4 Algorithm

7.4.1 The Loop Over Energy Groups

The scattering integral in the multigroup form of the Boltzmann equation, Eq. (7.3), contains a sum over all energy groups, $g' = 1, \ldots, G$. It can be seen, then, that to solve the equation for group g one needs to know the fluences for all the other energy groups. While this does not make the problem intractable, the algorithm can be simplified significantly if $\sigma_s(\vec{r}, \vec{\Omega} \cdot \vec{\Omega}', g' \rightarrow g) = 0$, for any $g < g'$. Because energy groups are numbered so that $E_g > E_{g'}$, if $g < g'$, this condition means that particles cannot gain energy in collisions. After a collision, the particle either stays in the same energy group (small or zero energy loss), or it transitions to a group of lower energy. The latter process is called down-scatter. There are, of course, some processes where the particle can gain energy, from a nuclear reaction or from an external electric field. These will not be considered here, as we aim to limit the discussion to the simplest algorithms.

Then, we can start with the equation for energy group 1. Because this group represents the highest energy, no particles will be down-scattered to this group, and the equation will not contain fluences for any energy groups other than group 1. This equation can be solved for fluence Φ_1, where the subscript indicates the energy group number ($g = 1$). Once we have found Φ_1, we can proceed to the equation for group 2. In this equation, the collision integral will include one down-scatter term representing particles scattering from group 1 to group 2. This term will contain the product $\sigma_s(1 \rightarrow 2)\Phi_1$, which is the collision density for scattering from group 1 down to group 2. The collision integral will also include the in-group scattering term, which has the form $\sigma_s(2 \rightarrow 2)\Phi_2$. Because the equation does not include fluences for any other groups, and at this step Φ_1 is already known, we can now solve the equation for fluence Φ_2. The next equation, for group 3, will have two down-scatter terms, $\sigma_s(1 \rightarrow 3)\Phi_1$ and $\sigma_s(2 \rightarrow 3)\Phi_2$, in addition to the in-group scattering term, $\sigma_s(3 \rightarrow 3)\Phi_3$. We can solve the equation for Φ_3, because at this step Φ_1 and Φ_2 are known. We then proceed to the next energy group, and so on, until we find the solution for the last energy group Φ_G. The flow chart of the process is shown in Fig. 7.2.

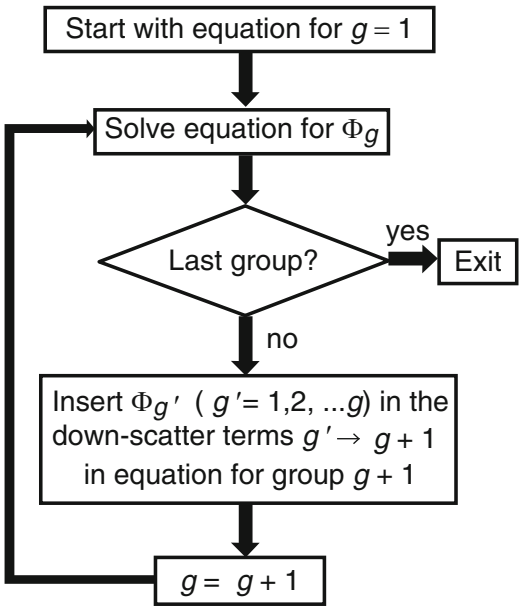

Fig. 7.2 The loop over energy groups. The *subscript* indicates the energy group number

7.4.2 Source Iterations

In the second box of the flow chart in Fig. 7.2, the Boltzmann equation is solved for the group fluence Φ_g. The solution is found through an iterative procedure called source iterations. The flow chart of this iterative algorithm is shown in Fig. 7.3. In the first step, the unscattered fluence Φ_0 is calculated. We discussed calculation of the unscattered fluence in Sect. 3.9.1. This step involves calculation of the optical distances between the source and the nodes of the tetrahedral mesh. After the unscattered fluence Φ_0 is found, it is inserted into the collision integral. In other words, the collision integral $\hat{K}\Phi$ is approximated by $\hat{K}\Phi_0$. Then the collision integral is added to the source function S on the right-hand side of the Boltzmann equation. Consequently, the right-hand side of the equation is known, and all the unknowns are in the left-hand side. In the discrete ordinates method (see Eqs. (7.28) and (7.29)), the left-hand side of the equation contains an unknown fluence only for one discrete direction, $\Phi(\vec{\Omega}_{ij})$. The fluences for all the other directions $\Phi(\vec{\Omega}_{i'j'})$ that were originally present on the right-hand side of the equation, because it uses the quadrature given by Eq. (7.29), have been calculated in the preceding source iteration step. This simplifies the algorithm substantially, because the equation can now be solved for one discrete direction at a time.

After a new solution is found (fluence Φ_{i+1} in box 3, where the subscript indicates the iteration number), it is inserted into the collision integral to replace the fluence from the preceding iteration, and then we need to solve the same equation with an updated right-hand side. These steps are repeated until a desired level of

accuracy is achieved for all nodes within a volume of interest. The accuracy can be estimated by comparing the results of a few, at least three, consecutive iterations to reduce the risk of false convergence.

7.5 Electron-Photon Fields

Electron-photon fields are particularly important for radiation therapy. To calculate a dose distribution from a photon or electron beam, we need the electron fluence, Φ_e. However, in a photon beam electrons are produced in photon collisions. Hence, to find the electron source, we need to know the photon fluence Φ_γ. On the other hand, in both photon and electron beams photons can be produced in electron collisions. To find this source of photons, we need to know the electron fluence, Φ_e. This means that we need to solve a system of two Boltzmann equations, one for the photon fluence and the other for the electron fluence. The equations are coupled, because the equation for Φ_γ includes the source of photons produced in electron collisions, the collisional source. This source has the following form (see, for example, Lorence et al. (1985)):

$$S_{e\gamma}\left(\vec{r},\vec{\Omega},E\right)=\int \mathrm{d}\vec{\Omega}'\int \mathrm{d}E'\,\sigma_{s,e\gamma}\left(\vec{r},\vec{\Omega}',E'\rightarrow\vec{\Omega},E\right)\Phi_e\left(\vec{r},\vec{\Omega}',E'\right), \tag{7.52}$$

where $\sigma_{s,e\gamma}$ is the double differential cross section for electron collisions producing photons. Similarly, the equation for Φ_e includes the source of electrons produced in photon collisions:

$$S_{\gamma e}\left(\vec{r},\vec{\Omega},E\right)=\int \mathrm{d}\vec{\Omega}'\int \mathrm{d}E'\,\sigma_{s,\gamma e}\left(\vec{r},\vec{\Omega}',E'\rightarrow\vec{\Omega},E\right)\Phi_\gamma\left(\vec{r},\vec{\Omega}',E'\right). \tag{7.53}$$

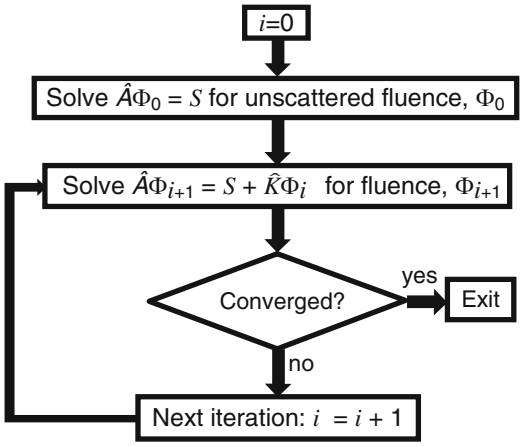

Fig. 7.3 Source iterations. The *subscript* indicates the iteration number

The system can be solved through an iterative procedure, such as source iterations on the collisional source, where at each step the equation for particles of type A is solved using the fluence of particles of type B found in the preceding iteration step inserted into the collisional source, S_{BA}. The partial coupling approximation is often used to avoid this extra iteration loop. In applications to electron-photon fields, the approximation can have one of two forms: (a) the production of electrons in photon collisions is neglected (Lorence et al. 1985), or, conversely, (b) the production of photons in electron collisions is neglected (Vassiliev et al. 2010). The former is not applicable to dose calculations in radiotherapy. It would result in unacceptable errors, especially near interfaces between different materials. The latter form works well in calculations of patient dose distributions as was shown by Vassiliev et al. (2010) and confirmed in a series of publications evaluating the accuracy of the Acuros® algorithm, for example, Bush et al. (2011), Hoffmann et al. (2012), Kan et al. (2012), to name a few. However, in problems involving high atomic number materials, photons produced in electron collisions, specifically, the Bremsstrahlung radiation, generally cannot be neglected. An important example where the partial coupling approximation, in either form, is not applicable is electron-photon transport in the beamline of a medical accelerator. If photons produced in electron collisions can be neglected, then the equation for the photon fluence can be solved separately, because in this case it does not include the electron fluence. Then, the photon fluence is inserted into the collisional source $S_{\gamma e}$ in the equation for the electron fluence, which can then be solved.

In problems where the photon energies exceed two electron rest energies, $E_\gamma > 2m_e c^2 \approx 1.022\,\text{MeV}$, a third type of particle, positrons, is produced in the pair production process. This, in principle, requires that a third Boltzmann equation, that for the positron fluence, be solved. This, however, can be avoided by replacing positrons with electrons. That is, two electrons are produced in a pair production event instead of an electron–positron pair. This is a good approximation, because electrons and positrons are similar in terms of energy loss and scattering parameters, and because the contribution to the dose from positrons is usually small.

The large interaction cross sections and very forward-peaked scattering of electrons pose serious difficulties for GBBS algorithms. As in the Monte Carlo method, achieving a high computing speed requires the use of multiple scattering theories for electron transport. In Acuros®, for example, soft collisions are treated in the continuous slowing down approximation (Vassiliev et al. 2010). We described the modification of the collision integral on which this method is based in Sect. 3.9.5. In Attila™, which is a more general purpose commercial software of this type, in addition to the continuous slowing down model for energy loss, the Fokker-Planck approximation (see Sect. 3.9.6) is applied to the angular variables to account for Coulomb scattering (Gifford et al. 2006; Vassiliev et al. 2008). These approximations introduce in the Boltzmann equation additional differential operators with derivatives in the energy and in angular variables.

7.6 Algorithm in an External Magnetic Field

The Boltzmann equation in the presence of an external magnetic field was derived in Sect. 3.8. It has the following form:

$$\left(\vec{f}_{\text{ex}}\left(\vec{p}\,\right)\cdot\vec{\nabla}_p\right)\left[\frac{1}{v}\Phi\left(\vec{r},\vec{p}\,\right)\right]+\left(\vec{\Omega}\cdot\vec{\nabla}_r\right)\Phi\left(\vec{r},\vec{p}\,\right)+\sigma\left(\vec{r},E\right)\Phi\left(\vec{r},\vec{p}\,\right)$$
$$=S\left(\vec{r},\vec{p}\,\right)+\int \mathrm{d}\vec{p}\,'\sigma_s\left(\vec{r},\vec{p}\,'\rightarrow\vec{p}\,\right)\Phi\left(\vec{r},\vec{p}\,'\right). \tag{7.54}$$

Here, $\vec{f}_{\text{ex}}$ is the force exerted on a charged particle by an external magnetic field. The force is given by the Lorentz formula (Jackson 1999)

$$\vec{f}_{\text{ex}}\left(\vec{p}\,\right)=q\,\vec{v}\times\vec{B}, \tag{7.55}$$

where q is the particle charge, $\vec{v}$ is its velocity, and $\vec{B}$ is the magnetic field. The force of an electric field will not be considered. The electric field $\vec{E}$ is set to zero.

We can now derive an algorithm for solving it. We will use an approach introduced by St. Aubin et al. (2015) and focus only on the first term in the equation, the term associated with the external force. We considered all the other terms of the equation earlier in this chapter.

7.6.1 The Magnetic Force Term in Spherical Coordinates

Before we can proceed further, we need the gradient operator $\vec{\nabla}_p\, g(\vec{p}\,)$ written in spherical coordinates, where $g(\vec{p}\,)$ is an arbitrary differentiable function. The spherical coordinates are defined as follows:

$$\begin{aligned} p_x &= p\sin\theta\cos\phi = p\sqrt{1-\mu^2}\cos\phi = p\,\Omega_x\,;\\ p_y &= p\sin\theta\sin\phi = p\sqrt{1-\mu^2}\sin\phi = p\,\Omega_y\,;\\ p_z &= p\cos\theta = p\mu = p\,\Omega_z\,. \end{aligned} \tag{7.56}$$

The gradient $\vec{\nabla}_p\, g(\vec{p}\,)$ is a vector:

$$\vec{\nabla}_p\, g=\vec{e}_x\frac{\partial g}{\partial p_x}+\vec{e}_y\frac{\partial g}{\partial p_y}+\vec{e}_z\frac{\partial g}{\partial p_z}. \tag{7.57}$$

where $\vec{e}_x$, $\vec{e}_y$, and $\vec{e}_z$ are the Cartesian unit basis vectors. Our goal is to express $\vec{\nabla}_p\, g(\vec{p}\,)$ in terms of the derivatives

$$\partial g/\partial p\,;\quad \partial g/\partial \mu\,;\quad \partial g/\partial \phi,$$

instead of the derivatives

$$\partial g/\partial p_x\,;\quad \partial g/\partial p_y\,;\quad \partial g/\partial p_z.$$

To achieve this, we use the chain rule:

$$\frac{\partial g}{\partial p_k} = \frac{\partial g}{\partial p}\frac{\partial p}{\partial p_k} + \frac{\partial g}{\partial \mu}\frac{\partial \mu}{\partial p_k} + \frac{\partial g}{\partial \phi}\frac{\partial \phi}{\partial p_k};\ \ k = x, y, z; \tag{7.58}$$

and find the derivatives

$$\partial p/\partial p_k\,;\quad \partial \mu/\partial p_k\,;\quad \partial \phi/\partial p_k\,;$$

from Eq. (7.56). The result is

$$\begin{aligned}\vec{\nabla}_p\, g = {} & \frac{\partial g}{\partial p}\left(\vec{e}_x\Omega_x + \vec{e}_y\Omega_y + \vec{e}_z\mu\right) \\ & + \frac{\partial g}{\partial \mu}\left(-\vec{e}_x\frac{\mu}{p}\Omega_x - \vec{e}_y\frac{\mu}{p}\Omega_y + \vec{e}_z\frac{1-\mu^2}{p}\right) \\ & + \frac{\partial g}{\partial \phi}\left(-\vec{e}_x\frac{\Omega_y}{p\,(1-\mu^2)} + \vec{e}_y\frac{\Omega_x}{p\,(1-\mu^2)}\right).\end{aligned} \tag{7.59}$$

This expression can be further simplified to

$$\vec{\nabla}_p g = \frac{\partial g}{\partial p}\vec{\Omega} - \frac{\partial g}{\partial \mu}\left(\frac{\mu}{p}\vec{\Omega} - \vec{e}_z\frac{1}{p}\right) - \frac{\partial g}{\partial \phi}\left(\vec{e}_x\frac{\Omega_y}{p\,(1-\mu^2)} - \vec{e}_y\frac{\Omega_x}{p\,(1-\mu^2)}\right). \tag{7.60}$$

Applying this result to the magnetic force term in the Boltzmann equation, Eq. (7.54), and using $([\vec{v}\times\vec{B}]\cdot\vec{\Omega}) = 0$, we derive the final form of the magnetic force term

$$\begin{aligned}&\left(\vec{f}_{\rm ex}\left(\vec{p}\,\right)\cdot\vec{\nabla}_p\right)\left[\frac{1}{v}\Phi\left(\vec{r},\vec{p}\,\right)\right] = q\left(\left[\vec{v}\times\vec{B}\right]\cdot\vec{\nabla}_p\right)\left[\frac{1}{v}\Phi\left(\vec{r},\vec{p}\,\right)\right] = \\ &\frac{qv}{p}\left[\vec{\Omega}\times\vec{B}\right]\cdot\left[\vec{e}_z\frac{\partial}{\partial\mu} - \left(\vec{e}_x\frac{\Omega_y}{(1-\mu^2)} - \vec{e}_y\frac{\Omega_x}{(1-\mu^2)}\right)\frac{\partial}{\partial\phi}\right]\left[\frac{1}{v}\Phi\left(\vec{r},\vec{p}\,\right)\right] = \\ &\frac{q}{p}\left[\vec{\Omega}\times\vec{B}\right]_z\frac{\partial}{\partial\mu}\Phi\left(\vec{r},\vec{p}\,\right) + \frac{q}{p\,(1-\mu^2)}\left[\vec{\Omega}\times\left[\vec{\Omega}\times\vec{B}\right]\right]_z\frac{\partial}{\partial\phi}\Phi\left(\vec{r},\vec{p}\,\right).\end{aligned} \tag{7.61}$$

7.6.2 Representation in Spherical Harmonics

The Boltzmann equation, Eq. (7.54), now has two derivatives with respect to the angular variables μ and ϕ. The equation can still be solved by the finite elements method applied this time to both the spatial coordinates x, y, z and the angular variables μ and ϕ. This approach to the problem was developed by St. Aubin et al. (2016) and produced excellent results. Here, however, we will present a simpler method that was proposed in an earlier paper (St. Aubin et al. 2015). This method is well suited for implementation in deterministic algorithms that use the discrete ordinates method.

The idea of the method is that the angular derivatives can be calculated analytically, if we expand the fluence in spherical harmonics, just as we did earlier in this chapter, in Eq. (7.10):

$$\begin{aligned}\frac{\partial}{\partial\mu}\Phi\left(\vec{r},\vec{\Omega}\right) &\approx \sum_{l=0}^{L}\sum_{m=-l}^{l}\Phi_l^m\left(\vec{r}\,\right)\frac{\partial}{\partial\mu}Y_l^m\left(\vec{\Omega}\right),\\ \frac{\partial}{\partial\phi}\Phi\left(\vec{r},\vec{\Omega}\right) &\approx \sum_{l=0}^{L}\sum_{m=-l}^{l}\Phi_l^m\left(\vec{r}\,\right)\frac{\partial}{\partial\phi}Y_l^m\left(\vec{\Omega}\right).\end{aligned} \tag{7.62}$$

The spherical harmonics are defined as follows (Arfken et al. 2013):

$$Y_l^m\left(\vec{\Omega}\right) = N_{l,m}P_l^m\left(\mu\right)\mathrm{e}^{im\phi} \tag{7.63}$$

where P_l^m is the associated Legendre polynomial and

$$N_{l,m} \equiv \sqrt{\frac{2l+1}{4\pi}\frac{(l-m)!}{(l+m)!}}. \tag{7.64}$$

These spherical harmonics are complex valued functions. St. Aubin et al. (2015), however, used real valued spherical harmonics that are more convenient for numerical implementation. Here we chose the complex valued functions for consistency with the rest of the book. Changing from complex valued to real valued spherical harmonics in the formalism that we present below is straightforward.

The first of the two angular derivatives, $\partial/\partial\mu$, can be found in Abramowitz and Stegan (1964):

$$\left(1-\mu^2\right)\frac{\mathrm{d}}{\mathrm{d}\mu}P_l^m\left(\mu\right) = \left(l+m\right)P_{l-1}^m\left(\mu\right) - \mu l P_l^m\left(\mu\right);\ l\geq 1,\ |m|\leq l-1. \tag{7.65}$$

For $l = m = 0$ the derivative is zero. For $m = \pm l$, we have (Arfken et al. 2013)

$$\begin{aligned}
\left(1-\mu^2\right)\frac{\mathrm{d}}{\mathrm{d}\mu}P_l^l(\mu) &= \left(1-\mu^2\right)\frac{\mathrm{d}}{\mathrm{d}\mu}(-1)^l(2l-1)!!\left(1-\mu^2\right)^{l/2} = -\mu l P_l^l(\mu)\,;\\
\left(1-\mu^2\right)\frac{\mathrm{d}}{\mathrm{d}\mu}P_l^{-l}(\mu) &= \left(1-\mu^2\right)\frac{\mathrm{d}}{\mathrm{d}\mu}\frac{(-1)^l}{(2l)!}P_l^l(\mu) = -\mu l P_l^{-l}(\mu)\,.
\end{aligned} \tag{7.66}$$

It is convenient to express the derivative in terms of spherical harmonics:

$$\begin{aligned}
\frac{\partial}{\partial\mu}Y_l^m\left(\vec{\Omega}\right) &= N_{l,m}\mathrm{e}^{im\phi}\frac{\mathrm{d}}{\mathrm{d}\mu}P_l^m(\mu)\\
&= \frac{1}{1-\mu^2}\left[\frac{N_{l,m}}{N_{l-1,m}}(l+m)\,Y_{l-1}^m\left(\vec{\Omega}\right) - \mu l Y_l^m\left(\vec{\Omega}\right)\right];\ \ l\geq 1;\ \ |m|\leq l-1.
\end{aligned} \tag{7.67}$$

For $|m| = l$, the result is

$$\frac{\partial}{\partial\mu}Y_l^{\pm l}\left(\vec{\Omega}\right) = N_{l,\pm l}\mathrm{e}^{\pm il\phi}\frac{\mathrm{d}}{\mathrm{d}\mu}P_l^{\pm l}(\mu) = \frac{-\mu l}{1-\mu^2}Y_l^{\pm l}\left(\vec{\Omega}\right). \tag{7.68}$$

The other angular derivative, $\partial/\partial\phi$, is simple

$$\frac{\partial}{\partial\phi}Y_l^m\left(\vec{\Omega}\right) = N_{l,m}P_l^m(\mu)\frac{\mathrm{d}}{\mathrm{d}\phi}\mathrm{e}^{im\phi} = imY_l^m\left(\vec{\Omega}\right). \tag{7.69}$$

We can now insert the derivatives of the spherical harmonics as given by Eqs. (7.67)–(7.69) into the sum in Eq. (7.62), and then insert the sums in the expression for the magnetic force term, Eq. (7.61). The result is the following rather lengthy formula:

$$\begin{aligned}
\left(\vec{f}_{\mathrm{ex}}\left(\vec{p}\,\right)\cdot\vec{\nabla}_p\right)\left[\frac{1}{v}\Phi\left(\vec{r},\vec{p}\,\right)\right] &\approx \frac{q}{p\left(1-\mu^2\right)}\left[\vec{\Omega}\times\vec{B}\right]_z\\
&\times\sum_{l=1}^{L}\left\{\sum_{m=-l+1}^{l-1}\Phi_l^m\left(\vec{r}\,\right)\left[\frac{N_{l,m}}{N_{l-1,m}}(l+m)\,Y_{l-1}^m\left(\vec{\Omega}\right)\right] - \sum_{m=-l}^{l}\mu l Y_l^m\left(\vec{\Omega}\right)\right\}\\
&+\frac{q}{p\left(1-\mu^2\right)}\left[\vec{\Omega}\times\left[\vec{\Omega}\times\vec{B}\right]\right]_z\sum_{l=0}^{L}\sum_{m=-l}^{l}im\,\Phi_l^m\left(\vec{r}\,\right)Y_l^m\left(\vec{\Omega}\right).
\end{aligned} \tag{7.70}$$

However, its overall structure is simple. It is the sum of the spherical harmonics multiplied by the expansion coefficients of the fluence, Φ_l^m, and by known coefficients that depend on the magnetic field strength $\vec{B}$ and direction $\vec{\Omega}$. We will leave this result as shown, without further discretization in space and energy. These steps are straightforward, and the general methodology was discussed earlier in this chapter.

The significance of the representation of the magnetic force term in spherical harmonics is that it brings the term to a form similar to that of the collision integral

in Eq. (7.19). Then, the magnetic force term can be formally considered as part of the collision integral and therefore the Boltzmann equation in the presence of a magnetic field can still be solved by an iterative procedure. However, St. Aubin et al. (2015) have shown that combining the magnetic force term given by Eq. (7.70) with the collision integral makes the source iterations algorithm unstable. For this reason instead of source iterations, the authors (St. Aubin et al. 2015) ultimately chose another method—a nonstationary Krylov subspace iterative technique GMRES (Generalized Minimal Residual Algorithm) (Krylov 1931; Saad and Schultz 1986).

For completeness, we include the definition of real-valued spherical harmonics. It differs from Eq. (7.63) only in that the exponent is replaced according to the following rule (Arfken et al. 2013):

$$e^{im\phi} \rightarrow \begin{cases} 1, & m = 0, \\ \sqrt{2}\cos(m\phi), & m > 0, \\ \sqrt{2}\sin(|m|\phi), & m < 0. \end{cases} \tag{7.71}$$

Comparison with Monte Carlo

This deterministic GBBS algorithm achieves the same level of accuracy as a full Monte Carlo algorithm, such as that implemented in the EGSnrc code system (Kawrakow et al. 2013).

St. Aubin et al. (2015) calculated dose distributions for 6 MV photon beams incident normally on a 30 cm cubic phantom comprising water, bone, and lung slabs, in magnetic fields of up to 3 T. They used the discrete ordinates (DO) GBBS algorithm described above and, for comparison, a full Monte Carlo code (Kawrakow et al. 2013). The DO GBBS results agreed with Monte Carlo calculations within 2 %/ 2 mm gamma criterion for 99.8 % of points, and within 1 %/ 1 mm for 94.8 % of points.

In Figs. 7.4 and 7.5 the depth dose on the central axis for a $10 \times 10\,\text{cm}^2$ calculated with the GBBS algorithm is compared with the Monte Carlo results. In both figures the magnetic field is 3 T. In Fig.7.4 the magnetic field is parallel to the beam. In this case the effect of the magnetic field on the dose distribution is relatively small. In contrast, in Fig.7.5 the magnetic field is perpendicular to the beam and the dose distribution is quite different, especially near the interfaces between different materials. For example, there is a distinct spike at the interface between bone and lung, at the depth of 12 cm. It was largely caused by electrons produced the bone, initially mostly in the forward direction, but forced by the magnetic field from the lung back towards the bone surface. This effect also reduces the fluence of electron reaching the water surface at 20 cm depth resulting in a pronounced dose minimum around this depth.

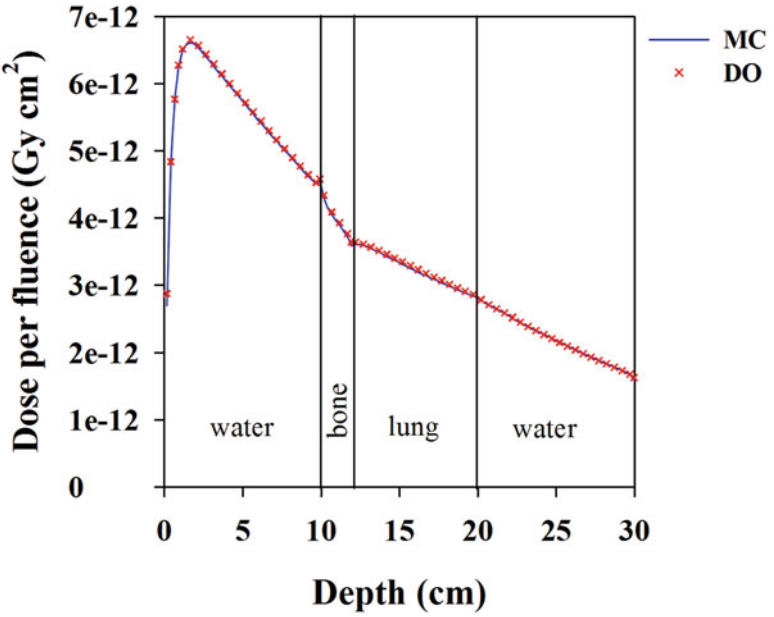

Fig. 7.4 Depth dose along the central axis for a $10 \times 10\,\text{cm}^2$ field, comparing Monte Carlo and the discrete ordinates GBBS formalism in a 3 T magnetic field parallel to the radiation beam [J. St. Aubin, private communication, 2016]. Used with permission

In that study limitations of the algorithm were also noted. In cases with low density materials in strong magnetic fields, the Krylov iterative algorithm converged too slowly, to a point where iterative stagnation occurred if the restart parameter was chosen too small. For example, in a lung phantom, iterations converged for magnetic field strengths of up to 6 T, and did not converge for magnetic fields stronger than 6 T. This limitation has been eliminated in a new algorithm (St. Aubin et al. 2016). This new algorithm does not use the discrete ordinates method. Instead, the authors developed a discontinuous finite-element space-angle approach to solving the Boltzmann equation with magnetic fields. The new algorithm has been validated against Geant4 Monte Carlo code (Allison et al. 2006), including dose calculations in an air cavity, and at bone–air interface. The overall agreement between the new GBBS algorithm and the Monte Carlo code was within 2 % / 2 mm gamma criterion for 98.9 % points.

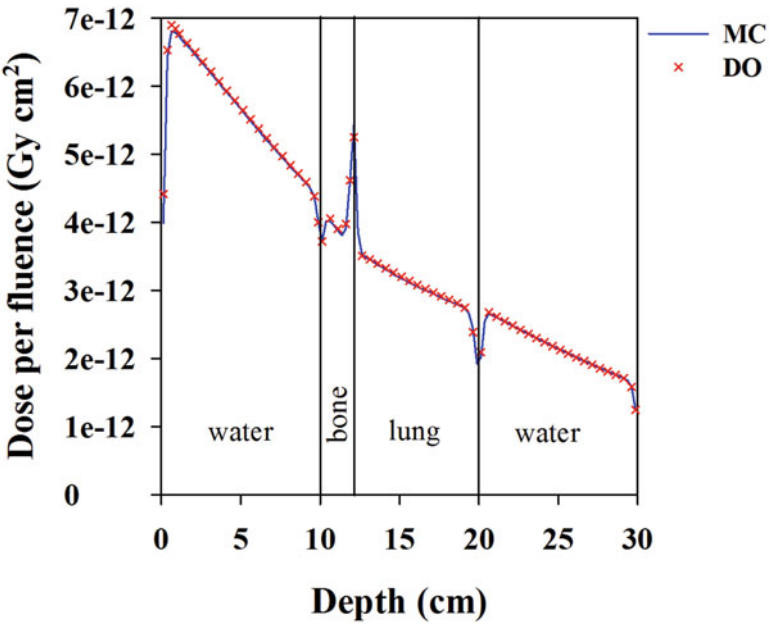

Fig. 7.5 Depth dose along the central axis for a $10 \times 10\,\text{cm}^2$ field comparing Monte Carlo and the discrete ordinates GBBS formalism in a 3 T magnetic field perpendicular to the radiation beam [J. St. Aubin, private communication, 2016]. Used with permission

References

Abramowitz, M., Stegan I.A. (eds.): Handbook of Mathematical Functions: With Formulas, Graphs, and Mathematical Tables. Dover Publications, New York (1964)

Allison, J., Amako, K., Apostolakis, J., Araujo, H., Arce Dubois, P., Asai, M., Barrand, G., Capra, R., et al.: Geant4 developments and applications. IEEE Trans. Nucl. Sci. **53**(1), 270–278 (2006)

Arfken, G.B., Weber, H.J., Harris F.E.: Mathematical Methods for Physicists, 7th edn. Elsevier, Amsterdam (2013)

Bush, K., Gagne, I.M., Zavgorodni, S., Ansbacher, W., Beckham, W.: Dosimetric validation of Acuros XB with Monte Carlo methods for photon dose calculations. Med. Phys. **38**(4), 2208–2221 (2011)

Dudewicz, E.J., Mishra S.N.: Modern Mathematical Statistics. Wiley, New York (1988)

Gifford, K.A., Horton J.L., Wareing T.A., Failla, G., Mourtada, F.: Comparison of a finite-element multigroup discrete-ordinates code with Monte Carlo for radiotherapy calculations. Phys. Med. Biol. **51**, 2253–2265 (2006)

Hoffmann, L., Jorgensen M.B.K., Muren, L.P., Petersen, J.B.B.: Clinical validation of the Acuros XB photon dose calculation algorithm, a grid-based Boltzmann equation solver. Acta Oncol. **51**(3), 376–385 (2012)

Jackson, J.D.: Classical Electrodynamics, 3rd edn. Wiley, New York (1999)

Kan, M.W.K., Leung, L.H.T., Yu, P.K.N.: Verification and dosimetric impact of Acuros XB algorithm on intensity modulated stereotactic radiotherapy for locally persistent nasopharyngeal carcinoma. Med. Phys. **39**(8), 4705–4714 (2012)

Kan, M.W.K., Leung, L.H.T., Yu, P.K.N.: Dosimetric impact of using the Acuros XB algorithm for intensity modulated radiation therapy and RapidArc planning in nasopharyngeal carcinomas. Int. J. Radiat. Oncol. Biol. Phys. **85**(1), E73–E80 (2013)

Kawrakow, I., Mainegra-Hing, E., Rogers, D.W.O., Tessier, F., Walters, B.R.B.: The EGSnrc Code System: Monte Carlo simulation of electron and photon transport. Report PIRS-701. National Research Council of Canada (2013)

Kroon, P.S., Hol, S., Essers, M.: Dosimetric accuracy and clinical quality of Acuros XB and AAA dose calculation algorithm for stereotactic and conventional lung volumetric modulated arc therapy plans. Radiat. Oncol. **8**(49), 149 (2013)

Krylov, A.N.: On the numerical solution of the equation by which in technical problems frequencies of small oscillations of material systems are determined. Izvestiya Academii Nauk SSSR **VII**(4), 491–539 (1931)

Lewis, E.E., Miller W.F.: Computational Methods of Neutron Transport. American Nuclear Society Inc., La Grange Park, IL (1993)

Lorence, L.J., Nelson W.E., Morel J.E.: Coupled electron-photon transport calculations using the method of discrete ordinates. IEEE Trans. Nucl. Sci. **32**(6), 4416–4419 (1985)

Saad, Y., Schultz, M.H.: GMRES: a generalized minimal residual algorithm for solving nonsymmetric linear systems. SIAM J. Sci. Stat. Comput. **7**(3), 865–869 (1986)

St. Aubin, J., Keyvanloo, A., Vassiliev, O., Fallone, B.G.: A deterministic solution of the first order linear Boltzmann transport equation in the presence of external magnetic fields. Med. Phys. **42**(2), 780–793 (2015)

St. Aubin, J., Keyvanloo, A., Fallone, B.G.: Discontinuous finite element space-angle treatment of the first order linear Boltzmann transport equation with magnetic fields: application to MRI-guided radiotherapy. Med. Phys. **43**(1), 195–204 (2016)

Tsuruta, Y., Nakata, M., Nakamura, M., Matsuo, Y., Higashimura, K., Monzen, H., Mizowaki, T., Hiraoka, M.: Dosimetric comparison of Acuros XB, AAA, and XVMC in stereotactic body radiotherapy for lung cancer. Med. Phys. **41**(8), 081715-1–081715-9 (2014)

Vassiliev, O.N., Wareing, T.A., Davis, I.M., McGhee, J., Barnett, D., Horton J.L., Gifford, K., Failla, G., Mourtada, F.: Feasibility of a multigroup deterministic solution method for three-dimensional radiotherapy dose calculations. Int. J. Radiat. Oncol. Biol. Phys. **72**(1), 220–227 (2008)

Vassiliev, O.N., Wareing, T.A., McGhee, J., Failla, G., Salehpour, M.R.: Validation of a new grid-based Boltzmann equation solver for dose calculation in radiotherapy with photon beams. Phys. Med. Biol. **55**(3), 581–598 (2010)

Walters, W.F.: Use of the Chebyshev-Legendre quadrature set in discrete-ordinate codes. Radiation Transport Group, Los Alamos National Laboratory (1987)

Appendix A
Probabilities and Statistics Refresher

A.1 Probability

Event A is the outcome of an experiment (trial). Sample space S is a set of all the possible outcomes of the experiment.

Example. If the trial is one roll of a die, then the sample space is $\{1, 2, 3, 4, 5, 6\}$.

The definition of the probability $P(A)$ of event A:

$$P(A) = \lim_{N \to \infty} \frac{n(A)}{N}, \tag{A.1}$$

where N is the number of trials and $n(A)$ is the number of times event A has occurred. The properties of probability:

1. $0 \leq P(A) \leq 1$.
2. $P(S) = 1$.
3. Additivity: if events A and B are mutually exclusive, that is $A \bigcap B = \emptyset$ (empty set), then $P(A \bigcup B) = P(A) + P(B)$.

A.2 Convergence in Probability

Definition (Dudewicz and Mishra 1988)**.** We say Z_n converges to Z in probability ($Z_n \xrightarrow{p} Z$) if for every $\epsilon > 0$

$$\lim_{n \to \infty} P\{|Z_n - Z| > \epsilon\} = 0.$$

Other convergence types are: (1) with probability one; (2) in mean square; (3) in distribution.

O.N. Vassiliev, *Monte Carlo Methods for Radiation Transport*, Biological and Medical Physics, Biomedical Engineering, DOI 10.1007/978-3-319-44141-2

A.3 Conditional Probability

The probability P of event A, given that event B has occurred:

$$P(A|B) = \frac{P(A \bigcap B)}{P(B)}. \tag{A.2}$$

A.4 Independent Events

Two events A and B are independent if $P(A|B) = P(A)$. If A and B are independent, then $P(A \bigcap B) = P(A) \cdot P(B)$.

A.5 The Total Probability Theorem

Let $\{A_i\}$, $i = 1, 2, \ldots, n$, be a complete set of mutually exclusive events. This means that $P\left(\bigcup_{i=1}^{n} A_i\right) = 1$ and $A_i \bigcap A_j = \emptyset$ for all $i \neq j$. Thus, in a trial one event from the set must occur, but any two events cannot occur. Then, for any event B:

$$P(B) = \sum_{i=1}^{n} P(A_i) P(B|A_i). \tag{A.3}$$

A.6 Probability Distribution of a Discrete Variable

If the sample space is countable, that is all possible outcomes can be enumerated: $A_1, A_2, \ldots$, then the probability distribution can be defined:

$$P(A_i) \equiv P(i) \equiv P_i; \quad i = 1, 2, \ldots \tag{A.4}$$

A.7 Probability Distribution Function

If the outcome of a trial is a number ξ, then the probability function $F_\xi(x)$ can be defined:

$$F_\xi(x) \equiv P\{\xi \leq x\}. \tag{A.5}$$

Random variable ξ may take a continuous range of values, a discrete set of values, or a combination of both. Variable x is continuous. Subscript ξ is often omitted.

The properties of the probability distribution function:

1.

$$F(x_2) \geq F(x_1), \text{ if } x_2 > x_1; \tag{A.6}$$

2.

$$\lim_{x \to \infty} F(x) = 1; \tag{A.7}$$

3.

$$F(x) \geq 0, \text{ for } \forall\, x. \tag{A.8}$$

A.8 Probability Density Function

If the derivative of the probability distribution function $F(x)$ exists, then the probability density function $f(x)$ can be defined:

$$f_\xi(x) \equiv \frac{\mathrm{d}F_\xi}{\mathrm{d}x}. \tag{A.9}$$

It defines the probability of observing the random variable within a given interval:

$$P\{\xi \in \mathrm{d}x\} = f_\xi(x)\,\mathrm{d}x. \tag{A.10}$$

$$P(a \leq \xi \leq b) = \int_a^b f_\xi(x)\,\mathrm{d}x. \tag{A.11}$$

A.9 Transformation of Random Variables and Their Distributions

One-Dimensional Random Variables Given: random variable ξ has a probability density $f_\xi(x)$; random variable η (the transformed variable) is a known function of ξ, $\eta = g(\xi)$. Problem: find the probability density of η. Solution:

$$f_\eta(y) = f_\xi(x)\left|\frac{\mathrm{d}x}{\mathrm{d}y}\right| = f_\xi(x)\left|\frac{\mathrm{d}g}{\mathrm{d}x}\right|^{-1}. \tag{A.12}$$

The result should be written using y rather than x. This can be done through the substitution $x = g^{-1}(y)$.

Example.

$$f_{\xi}(x) = 2x \exp\left(-x^2\right), \quad x \geq 0. \tag{A.13}$$

Find the distribution $f_{\eta}(y)$, where $\eta = \xi^2$. Solution

$$f_{\eta}(y) = 2x \exp\left(-x^2\right) |2x|^{-1} = \exp(-y). \tag{A.14}$$

Two-Dimensional Random Variables Given: two random variables (ξ_1, ξ_2) that have a joint probability density $f_{\xi_1,\xi_2}(x_1, x_2)$; two random variables (η_1, η_2) (the transformed variables) are known functions of (ξ_1, ξ_2), $\eta_1 = g_1(\xi_1, \xi_2)$ and $\eta_2 = g_2(\xi_1, \xi_2)$. Problem: find the joint probability density of (η_1, η_2). Solution:

$$f_{\eta_1,\eta_2}(y_1, y_2) = f_{\xi_1,\xi_2}(x_1, x_2) |J|, \tag{A.15}$$

where J is the Jacobian:

$$J = \begin{vmatrix} \partial x_1/\partial y_1 & \partial x_1/\partial y_2 \\ \partial x_2/\partial y_1 & \partial x_2/\partial y_2 \end{vmatrix}. \tag{A.16}$$

To find the derivatives, we need to find the inverse transformation $\xi_1 = g_1^{-1}(\eta_1, \eta_2)$ and $\xi_2 = g_2^{-1}(\eta_1, \eta_2)$. Then

$$\frac{\partial x_i}{\partial y_j} = \frac{\partial g_i^{-1}}{\partial y_j}; \quad i, j = 1, 2. \tag{A.17}$$

The result should be written using (y_1, y_2) rather than (x_1, x_2). This can be done through the substitutions $x_1 = g_1^{-1}(y_1, y_2)$ and $x_2 = g_2^{-1}(y_1, y_2)$.

Example. Transformation from Cartesian coordinates (x, y) to polar coordinates (ρ, ϕ). The inverse transformation: $x = \rho \cos\phi$, $y = \rho \sin\phi$. First, we find the Jacobian

$$J = \begin{vmatrix} \partial x/\partial \rho & \partial x/\partial \phi \\ \partial y/\partial \rho & \partial y/\partial \phi \end{vmatrix} = \rho. \tag{A.18}$$

Then, we can find the distribution:

$$f(\rho, \phi) = f(x(\rho, \phi), y(\rho, \phi)) \rho. \tag{A.19}$$

For example, if $f(x, y)$ is a uniform distribution within a circle of radius R, that is

$$f(x, y) = \frac{1}{\pi R^2}, \quad \text{if } x^2 + y^2 < R^2. \tag{A.20}$$

then,

$$f(\rho,\phi)=\frac{\rho}{\pi R^2},\ \text{ if }\ \rho<R. \tag{A.21}$$

Three-Dimensional Random Variables Given: three random variables (ξ_1, ξ_2, ξ_3) that have a joint probability density $f_{\xi_1,\xi_2,\xi_3}(x_1,x_2,x_3)$; three random variables (η_1,η_2,η_3) (the transformed variables) are known functions of (ξ_1,ξ_2,ξ_3), $\eta_i = g_i(\xi_1,\xi_2,\xi_3)$, $i=1,2,3$. Problem: find the joint probability density of (η_1,η_2,η_3). The solution is similar to that in the two-dimensional case, except for the Jacobian that is

$$J=\begin{vmatrix}\partial x_1/\partial y_1 & \partial x_1/\partial y_2 & \partial x_1/\partial y_3\\ \partial x_2/\partial y_1 & \partial x_2/\partial y_2 & \partial x_2/\partial y_3\\ \partial x_3/\partial y_1 & \partial x_3/\partial y_2 & \partial x_3/\partial y_3\end{vmatrix}. \tag{A.22}$$

Example. Transformation from Cartesian coordinates (x,y,z) to spherical coordinates (ρ,θ,ϕ). The inverse transformation: $x = r\sin\theta\cos\phi$, $y = r\sin\theta\sin\phi$, $z = r\cos\theta$. First, we find the Jacobian

$$J=\begin{vmatrix}\partial x/\partial r & \partial x/\partial\theta & \partial x/\partial\phi\\ \partial y/\partial r & \partial y/\partial\theta & \partial y/\partial\phi\\ \partial z/\partial r & \partial z/\partial\theta & \partial z/\partial\phi\end{vmatrix}=r^2\sin\theta. \tag{A.23}$$

Then, we can find the distribution:

$$f(r,\theta,\phi)=f\left(x(r,\theta,\phi),y(r,\theta,\phi),z(r,\theta,\phi)\right)r^2\sin\theta. \tag{A.24}$$

For example, if $f(x,y,z)$ is a uniform distribution within a sphere of radius R, that is

$$f(x,y,z)=\frac{3}{4\pi R^3},\ \text{ if }\ x^2+y^2+z^2<R^2, \tag{A.25}$$

then,

$$f(r,\theta,\phi)=\frac{3r^2\sin\theta}{4\pi R^3},\ \text{ if }\ r<R. \tag{A.26}$$

The changing from θ to $\mu=\cos\theta$, simplifies the result above

$$f(r,\mu,\phi)=\frac{3r^2}{4\pi R^3},\ \text{ if }\ r<R. \tag{A.27}$$

A.10 Distribution of a Sum

If ξ and η are statistically independent random variables with distributions $f_\xi(x)$ and $f_\eta(x)$, then the distribution of the sum $\zeta = \xi + \eta$ is given by

$$f_\zeta(x) = \int_{-\infty}^{\infty} f_\xi\left(x'\right) f_\eta\left(x - x'\right) \mathrm{d}x' \,. \tag{A.28}$$

The integral above is called the convolution and is denoted as $(f_\xi * f_\eta)(x)$.

A.11 The Expectation Value

For a continuous random variable ξ the definition is:

$$\mathrm{E}\{\xi\} \equiv \int_{-\infty}^{+\infty} x f(x)\, \mathrm{d}x; \tag{A.29}$$

For a discrete variable:

$$\mathrm{E}\{\xi\} \equiv \sum_k k P_k. \tag{A.30}$$

The expectation value is also referred to as the mean value.

The properties of the expectation value:

$$\mathrm{E}\{\xi_1 + \xi_2\} = \mathrm{E}\{\xi_1\} + \mathrm{E}\{\xi_2\}. \tag{A.31}$$

$$\mathrm{E}\{C\xi\} = C\mathrm{E}\{\xi\}; \text{ where } C \text{ is a constant.} \tag{A.32}$$

A.12 The Variance

The definition is:

$$\mathrm{Var}(\xi) \equiv \mathrm{E}\left(\xi^2\right) - [\mathrm{E}(\xi)]^2 = \mathrm{E}\{\xi - \mathrm{E}(\xi)\}^2; \tag{A.33}$$

It is a measure of the dispersion of a random variable.

The properties of the variance:

$$\mathrm{Var}(\xi_1 + \xi_2) = \mathrm{Var}(\xi_1) + \mathrm{Var}(\xi_2)\,; \text{ only if } \xi_1 \text{ and } \xi_2 \text{ are independent.} \tag{A.34}$$

$$\mathrm{Var}(C\xi) = C^2 \mathrm{Var}(\xi)\,; \text{ where } C \text{ is a constant.} \tag{A.35}$$

$$\mathrm{Var}(C) = 0. \tag{A.36}$$

The standard deviation:

$$\sigma \equiv \sqrt{\mathrm{Var}\,(\xi)}. \tag{A.37}$$

Chebyshev's theorem: for any real k

$$P\{|\xi - \mathrm{E}\,(\xi)\,| \geq k\sigma\} \leq \frac{1}{k^2}. \tag{A.38}$$

For example, if $k = 2$:

$$P\{|\xi - \mathrm{E}\,(\xi)\,| \geq 2\sigma\} \leq \frac{1}{4}. \tag{A.39}$$

A.13 The Normal Distribution

$$f_\xi\,(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[\frac{-(x-\mu)^2}{2\sigma^2}\right]. \tag{A.40}$$

$$\mathrm{E}\,(\xi) = \mu; \quad \mathrm{Var}\,(\xi) = \sigma^2. \tag{A.41}$$

$$P\{\mu - \sigma < \xi < \mu + \sigma\} \approx 0.683. \tag{A.42}$$

$$P\{\mu - 2\sigma < \xi < \mu + 2\sigma\} \approx 0.954. \tag{A.43}$$

A.14 The Central Limit Theorem

If $\xi_1, \xi_2, \ldots, \xi_N$ are independent random numbers and each has an arbitrary distribution with mean μ and finite variance σ^2, and

$$\eta = \frac{1}{N}\sum_{i=1}^{N} \xi_i, \tag{A.44}$$

then in the limit $N \to \infty$, the distribution $f_\eta\,(x)$ tends to a normal distribution with $\mu_\eta = \mu$ and $\sigma_\eta = \sigma/\sqrt{N}$.

A.15 The Binomial Distribution

Bernoulli trials:

1. Trials are independent events.
2. Each trial results in exactly one of the two mutually exclusive outcomes, called success and failure.
3. The probability of success is constant from trial to trial.

The Bernoulli process, or the Bernoulli sequence, is a sequence of Bernoulli trials. The binomial distribution is the distribution of number of successes k in n Bernoulli trials:

$$P(k|n) = \binom{n}{k} p^k (1-p)^{n-k}; \tag{A.45}$$

where $0 \leq k \leq n$, p is the probability of "success" in one trial, and

$$\binom{n}{k} = \frac{n!}{k!\,(n-k)!} \tag{A.46}$$

is the binomial coefficient.

A.16 The Poisson Distribution

It is derived from the binomial distribution by taking the limit $n \to \infty$, keeping the average number of successes, np, constant.

$$P_k = \frac{\mu^k}{k!} \mathrm{e}^{-\mu}; \;\; k = 0, 1, 2 \ldots \tag{A.47}$$

where μ is a parameter of the distribution, $\mu = \mathrm{E}(k) = \mathrm{Var}(k)$.

Example. The probability of k radioactive decays in 1 s is given by Eq. (A.47). In this case μ is the average number of decays per second (= activity in Bq).

A.17 The χ^2 Distribution

If ξ_i $(i = 1, 2, \ldots, r)$ have normal distributions with the mean equal to 0, and the variance equal to 1, then the sum

$$\chi^2 = \sum_{i=1}^{r} \xi_i^2 \tag{A.48}$$

is χ^2 distributed. Parameter of the distribution r is called the number of degrees of freedom.

$$f_{\chi^2}(x) = \frac{x^{r/2-1}\exp(-x/2)}{\Gamma(r/2)\,2^{r/2}}. \tag{A.49}$$

$$\mathrm{E}\left(\chi^2\right) = r. \tag{A.50}$$

$$\mathrm{Var}\left(\chi^2\right) = 2r. \tag{A.51}$$

A.18 Other Characteristics of a Distribution

The most probable value. It corresponds to the maximum of the distribution. It is often, but not always, can be found by solving:

$$\frac{\partial f}{\partial x} = 0. \tag{A.52}$$

The median, $x_{1/2}$.

$$\frac{1}{2} = \int_{-\infty}^{x_{1/2}} f(x)\,\mathrm{d}x; \tag{A.53}$$

The lower, $x_{1/4}$, and upper, $x_{3/4}$, quartiles.

$$\frac{1}{4} = \int_{-\infty}^{x_{1/4}} f(x)\,\mathrm{d}x; \quad \frac{3}{4} = \int_{-\infty}^{x_{3/4}} f(x)\,\mathrm{d}x. \tag{A.54}$$

The quantile, x_q.

$$q = \int_{-\infty}^{x_q} f(x)\,\mathrm{d}x; \quad 0 < q < 1. \tag{A.55}$$

The interquartile range (IQR) = $x_{3/4} - x_{1/4}$.

A.19 Statistics

The statistic is a quantity, which is calculated from observed data. In the following examples: $\{\xi_1, \xi_2, \ldots \xi_N\}$ is the observed data, or the sample, and N is the sample size.

The sample average.

$$\bar{\xi} = \frac{1}{N} \sum_{i=1}^{N} \xi_i. \tag{A.56}$$

The sample median.

$$\xi_{1/2} = \begin{cases} \xi_{\frac{N+1}{2}}, & \text{if } N \text{ is odd} \\ \\ \frac{1}{2}\left(\xi_{\frac{N}{2}} + \xi_{\frac{N}{2}+1}\right), & \text{if } N \text{ is even.} \end{cases} \tag{A.57}$$

The sample in Eq. (A.57) is assumed to be sorted in the increasing order: $\xi_1 \leq \xi_2 \leq \cdots \leq \xi_N$.

Sample quartiles and quantiles. The sample median as well as sample quartiles and quantiles are examples of order statistics. Order statistics are defined by the location of a random value in a sorted sample.

The sample variance, unbiased.

$$s^2 = \frac{1}{N-1} \sum_{i=1}^{N} \left(\xi_i - \bar{\xi}\right)^2. \tag{A.58}$$

A.20 Point Estimators

Hypothesis: ξ has a distribution $f_\xi(x, a, b, \ldots)$, where $a, b, \ldots$ are unknown parameters.

Problem. Given: Sample $\{\xi_1, \xi_2 \ldots \xi_N\}$. Estimate $a, b, \ldots$.

For each parameter $a, b, \ldots$ an estimator that is a function of the sample $\hat{\Theta}_a(x_1, x_2, \ldots x_N)$, etc., should be designed. Good estimators are:

1. Unbiased, $\mathrm{E}\{\hat{\Theta}_a(\xi_1, \xi_2, \ldots \xi_N)\} = a$.
2. Efficient, $\mathrm{Var}\{\hat{\Theta}_a\}$ is minimal.
3. Robust. Unusually small or large values in the sample have limited influence on the result. Other characteristics of robustness also exist, for example, see Hampel (1986).

A.21 The Maximum Likelihood Method

It is a method for deriving estimators that is based on the assumption that the acquired sample is the most probable. For example, if the probability distribution has two parameters, $f(x, a, b)$, then:

1. The likelihood function is written:

$$L(a, b) = \prod_{i=1}^{N} f(\xi_i, a, b). \tag{A.59}$$

2. The derivatives are calculated and a system of two equations is written

$$\frac{\partial L}{\partial a} = 0; \quad \frac{\partial L}{\partial b} = 0. \tag{A.60}$$

3. The system is solved for a and b. Then the expression for a is the estimator for a, $\hat{\Theta}_a(\xi_1, \xi_2, \ldots \xi_N)$. Similarly, the expression for b is $\hat{\Theta}_b$.

The algebra is often simpler if $\ln L$ is used instead of L. Moreover, the derivative equal to 0 does not always give a maximum. Sometimes the derivative does not even exist. The main advantage of this approach is that the most likelihood estimators are asymptotically efficient.

Example. Estimate parameter μ of the normal distribution.

First, we write the likelihood function:

$$\ln L(\mu, \sigma) = \sum_{i=1}^{N} \ln f(\xi_i, \mu, \sigma^2), \tag{A.61}$$

where $f(\xi_i, \mu, \sigma^2)$ is a normal distribution with parameters μ, σ^2. Then we solve the equation

$$\frac{\partial \ln L}{\partial \mu} = 0 \tag{A.62}$$

for μ. The solution is:

$$\mu = \frac{1}{N} \sum_{i=1}^{N} \xi_i; \tag{A.63}$$

Then, estimator for μ is the sample average

$$\hat{\Theta}_\mu = \frac{1}{N} \sum_{i=1}^{N} \xi_i. \tag{A.64}$$

A.22 The Method of Moments

This is another method for deriving point estimators. It is a simpler alternative to the most likelihood method. For example, if two parameters a and b need to be estimated, then:

1. The two equations, for the first and second moments are written:
   the sample mean = expectation (the first moment);
   the sample variance = the variance of the distribution (the second moment).
2. The aforementioned system of equations is solved for a and b.
3. The expression for a is the estimator Θ_a. Similarly, the estimator for b is derived.

Example. Estimate parameter μ of the normal distribution. The answer is

$$\hat{\Theta}_\mu = \frac{1}{N}\sum_{i=1}^{N}\xi_i. \tag{A.65}$$

A.23 Order Statistics as Estimators

The approach is based on the following property: as the sample size increases, the sample quantile ξ_q tends to the respective quantile of the distribution, x_q.

If, for example, two parameters of a distribution, a and b, need to be estimated, then:

1. Two quantile levels q_1 and q_2 are selected; for example, q_1=1/4 and $q_2 = 3/4$.
2. Two equations are written

$$\xi_{q1} = x_{q1}(a,b); \quad \xi_{q2} = x_{q2}(a,b). \tag{A.66}$$

3. The aforementioned system of the two equations is solved for a and b.
4. The expression for a is the estimator $\hat{\Theta}_a$. Similarly the estimator for b is derived.

Example. Estimate parameter μ of the normal distribution. The answer is

$$\hat{\Theta}_\mu = \xi_{1/2}. \tag{A.67}$$

To derive the estimator above, quantile level $q = 1/2$ was chosen.

A.24 Interval Estimates

The point estimator $\hat{\Theta}_a$ of parameter a produces an estimate that is likely to be close to the true value of a. The purpose of interval estimation is to find an interval that contains the true value of the parameter with a given probability p, for example $p = 0.95$.

Let random variable ξ have a distribution $f(x, a)$, where a is an unknown parameter, and we have designed a point estimator $\hat{\Theta}_a$. Next, we need to find the distribution $g(x, a)$ of the estimate. The distribution is a function of still unknown parameter a and the sample size N. The latter is omitted for brevity. For example, if $\hat{\Theta}_a$ is the sample average and the sample size N is sufficiently large, g can be approximated by a normal distribution.

To find the lower, L, and upper, U, boundaries of the interval, we define two functions $q_1(a)$, and $q_2(a)$ by the following equations:

$$\int_{-\infty}^{q_1(a)} g(x, a)\mathrm{d}x = (1-p)w\,; \tag{A.68}$$

$$\int_{q_2(a)}^{\infty} g\,(x, a)\,\mathrm{d}x = (1-p)(1-w)\,, \tag{A.69}$$

where w is any number satisfying $0 < w < 1$. Parameter w is present because for a given p multiple confidence intervals may exist. Then, the interval boundaries are calculated as the inverse of the functions above:

$$L = q_2^{-1}\left(\hat{\Theta}_a\right); \tag{A.70}$$

$$U = q_1^{-1}\left(\hat{\Theta}_a\right), \tag{A.71}$$

where $\hat{\Theta}_a$ is the "observed" estimate, that is $\hat{\Theta}_a(\xi_1, \xi_2, \ldots, \xi_N)$.

Example. $f(x, a)$ is a normal distribution with parameters μ and σ. We denote this distribution as $N(\mu, \sigma^2)$. The unknown parameter is $a = \mu$, and $\hat{\Theta}_\mu$ is the sample average. For simplicity, we assume that σ is known. We choose $p = 0.95$ and $w = 1/2$. Distribution $g(x, \mu)$ is also normal, $N(\mu, \sigma_N^2)$, where $\sigma_N^2 = \sigma^2/N$.

Then, Eqs. (A.68) and (A.69) lead to

$$\mathrm{erf}\left(\frac{q_1 - \mu}{\sigma_N\sqrt{2}}\right) = -0.95, \tag{A.72}$$

$$\mathrm{erf}\left(\frac{q_2 - \mu}{\sigma_N\sqrt{2}}\right) = 0.95, \tag{A.73}$$

resulting in

$$q_1\,(\mu) = \mu - 1.96\,\sigma_N, \tag{A.74}$$

$$q_2\,(\mu) = \mu + 1.96\,\sigma_N. \tag{A.75}$$

Finally, using Eqs. (A.70) and (A.71), we find the lower and upper boundaries of the interval:

$$L = \hat{\Theta}_\mu - 1.96\,\sigma_N; \tag{A.76}$$

$$U = \hat{\Theta}_\mu + 1.96\,\sigma_N\,. \tag{A.77}$$

A.25 Markov Process

A random process is a sequence of random numbers, or states: $\xi_1, \xi_2, \ldots$ A Markov process is a random process in which the future is independent of the past, given the present:

$$P\{\xi_n \mid \xi_1,\ \xi_2, \ldots \xi_{n-1}\} = P\{\xi_n \mid \xi_{n-1}\}; \tag{A.78}$$

that is, the conditional distribution of ξ_n given $\xi_1, \xi_2 \ldots, \xi_{n-1}$ equals to the conditional distribution of ξ_n given only ξ_{n-1}. Here, state ξ_{n-1} is "the present" and state ξ_n is "the future."

A Markov chain is defined by the probability distribution of the initial state, $\pi(x_0)$, and the transition probability $p\,(x \to x')$.

Note:

$$\int p\,(x \to x')\mathrm{d}x' \equiv 0 \leq p_s(x) \leq 1; \quad \text{for } \forall x, \tag{A.79}$$

where p_s is called the survival probability.

Appendix B
Useful Mathematics

B.1 The Gauss-Ostrogradsky Theorem

The theorem is also known as the divergence theorem. First, we introduce the notations. For vector field $\vec{A}\,(x, y, z)$:

$$\mathrm{div}\vec{A} = \left(\vec{\nabla}\cdot\vec{A}\right) = \frac{\partial A_x}{\partial x} + \frac{\partial A_y}{\partial y} + \frac{\partial A_z}{\partial z}\,, \tag{B.1}$$

where $\vec{\nabla}$ is a vector with three components:

$$\vec{\nabla} = \left\{\frac{\partial}{\partial x}, \frac{\partial}{\partial y}, \frac{\partial}{\partial z}\right\}. \tag{B.2}$$

Theorem. *For volume V with surface Γ and outward pointing normal to surface $\vec{n}_\Gamma$:*

$$\int_V \left(\vec{\nabla}\cdot\vec{A}\right)\mathrm{d}V = \oint_\Gamma \left(\vec{A}\cdot\vec{n}_\Gamma\right)\mathrm{d}\Gamma\,. \tag{B.3}$$

B.2 Transforming a Line Integral into a Volume Integral

Here we prove the following identity:

$$\int f\left(\vec{r}\,'\right)\delta\left(\vec{\Omega} - \frac{\vec{r}-\vec{r}\,'}{|\vec{r}-\vec{r}\,'|}\right)\frac{\mathrm{d}\vec{r}\,'}{|\vec{r}-\vec{r}\,'|^2} = \int_0^\infty f\left(\vec{r} - t\vec{\Omega}\right)\mathrm{d}t. \tag{B.4}$$

O.N. Vassiliev, *Monte Carlo Methods for Radiation Transport*, Biological and Medical Physics, Biomedical Engineering, DOI 10.1007/978-3-319-44141-2

In the first integral we introduce a new integration variable $\vec{R} = \vec{r} - \vec{r}\,'$. The integral becomes:

$$\int f\left(\vec{r} - \vec{R}\right) \delta\left(\vec{\Omega} - \frac{\vec{R}}{R}\right) \frac{\mathrm{d}\vec{R}}{R^2}. \tag{B.5}$$

Then we switch to the spherical coordinates:

$$\mathrm{d}\vec{R} = R^2 \mathrm{d}R \mathrm{d}\vec{\Omega}\,'; \quad \vec{R} = R\vec{\Omega}\,', \tag{B.6}$$

where $\mathrm{d}\vec{\Omega}\,'$ is the solid angle corresponding to the direction defined by unit vector $\vec{\Omega}\,'$. As a result, we obtain

$$\int_0^\infty \mathrm{d}R \int \mathrm{d}\vec{\Omega}\,' f\left(\vec{r} - R\vec{\Omega}\,'\right) \delta\left(\vec{\Omega} - \vec{\Omega}\,'\right). \tag{B.7}$$

The last step is the integration over $\vec{\Omega}\,'$ resulting in

$$\int_0^\infty f\left(\vec{r} - R\vec{\Omega}\right) \mathrm{d}R. \tag{B.8}$$

The denotation of integration variable t as R proves the identity.

□

The proof for the integration over an infinite volume was given. The identity is also applicable to any finite volume. In this case, the integration over t is from 0 to $t_{\max}(\vec{\Omega})$, which is the distance to the volume boundary from point $\vec{r}$ along a straight line, in the direction opposite to $\vec{\Omega}$.

B.3 Adjoint Operators

Definition. Operator $\hat{A}^+$ is adjoint to operator $\hat{A}$ if

$$\left(g, \hat{A}f\right) = \left(f, \hat{A}^+ g\right). \tag{B.9}$$

Example 1. Linear integral operator with kernel $k\,(x' \to x)$:

$$\hat{K}f = \int k\left(x' \to x\right) f\left(x'\right) \mathrm{d}x\,; \tag{B.10}$$

We need to find adjoint operator $\hat{K}^+$.

$$\begin{aligned}\left(g, \hat{K}f\right) &= \int \mathrm{d}x\, g\,(x) \int \mathrm{d}x'\, k\left(x' \to x\right) f\left(x'\right) \\ &= |\text{ change notations : } x = x'\,;\; x' = x\,| \\ &= \int \mathrm{d}x'\, g\left(x'\right) \int \mathrm{d}x\, k\left(x \to x'\right) f\,(x) \\ &= \int \mathrm{d}x f\,(x) \int \mathrm{d}x'\, k\left(x \to x'\right) g\left(x'\right) = \left(f, \hat{K}^{+} g\right)\,, \end{aligned} \tag{B.11}$$

where

$$\hat{K}^{+} g = \int \mathrm{d}x'\, k\left(x \to x'\right) g\left(x'\right)\,. \tag{B.12}$$

Conclusion: $\hat{K}^{+}$ is the operator with the transposed kernel, $k\,(x \to x')$.

Example 2. Matrix operator: the product of a matrix and a vector. In this case, it is easy to show that the adjoint operator is the multiplication by the transposed matrix.

Example 3. The derivative $(\mathrm{d}/\mathrm{d}x)$. Let us find the adjoint operator, $(\mathrm{d}/\mathrm{d}x)^{+}$.

$$\left(g, \frac{\mathrm{d}f}{\mathrm{d}x}\right) = \int_{-\infty}^{+\infty} g\,(x)\, \frac{\mathrm{d}f\,(x)}{\mathrm{d}x} \mathrm{d}x = |\text{ integrate by parts }| =$$

$$= g\,(x) f\,(x)\,|_{-\infty}^{+\infty} - \int_{-\infty}^{+\infty} f\,(x)\, \frac{\mathrm{d}g\,(x)}{\mathrm{d}x} \mathrm{d}x \;= g\,(x) f\,(x)\,|_{-\infty}^{+\infty} - \left(f, \frac{\mathrm{d}g}{\mathrm{d}x}\right). \tag{B.13}$$

Here, we need to add the condition that functions f and g are zeros at $-\infty$ and at $+\infty$. In this case:

$$(\mathrm{d}/\mathrm{d}x)^{+} = -\,(\mathrm{d}/\mathrm{d}x)\,. \tag{B.14}$$

Example 4. A sum of operators. Obviously:

$$\left(\hat{A} + \hat{B}\right)^{+} = \hat{A}^{+} + \hat{B}^{+}\,. \tag{B.15}$$

References

Dudewicz, E.J., Mishra S.N.: Modern Mathematical Statistics. Wiley, New York (1988)
Hampel, F.R.: Robust Statistics: The Approach Based on Influence Functions. Wiley, New York, 1986

List of Notations

A, a, α

A	(1) Area	Eq.(3.16)
	(2) normalization constant	Eq. (2.33)
a	Screening length	Eq. (5.217)

B, b, β

B	Parameter of Molière distribution	Eq. (5.108)
$B(z)$	Biological response function	Eq. (6.15)
$\vec{B}$	Magnetic field strength	Eq. (3.78)
β	Stopping power	Eq. (3.116)
$\beta_<$	Restricted stopping power	Eq. (3.130)
$\beta_>$	Stopping power for hard collisions	Eq. (5.52)

C, c

$\mathrm{Ci}(x)$	Cosine integral	Eq. (5.57)
c	Speed of light	Eq. (5.33)

D, d, Δ, δ

D	Detector response function	Eq. (3.28)
$\tilde{D}$	Modified detector response function	Eq. (4.61)
d	Width of charge state distribution	Eq. (5.213)
$d(y)$	Dose-weighted single event distribution of lineal energy	Eq. (6.11)
$d(z)$	Dose-weighted single event distribution of specific energy	Eq. (6.12)

O.N. Vassiliev, *Monte Carlo Methods for Radiation Transport*, Biological and Medical Physics, Biomedical Engineering, DOI 10.1007/978-3-319-44141-2

ΔE_g	Width of energy group g	p. 226
$(\Delta E)_{\max}$	Maximum energy transfer in one collision	Eq. (5.33)
Δx_i	Width of a histogram bin	Eq. (2.81)
$\delta(x)$	Delta function	Eq. (3.97)
δ_{ij}	Kronecker delta	Eq. (3.151)

E, e, ϵ

E	Energy	p. 49
$\mathrm{E}\{\cdot\}$	Expectation value	Eq. (A.29)
E_g	Lower boundary of energy group g	p. 226
$\mathrm{Ei}(x)$	Exponential integral	Eq. (5.40)
$\vec{e}_x, \vec{e}_y, \vec{e}_z$	Cartesian basis vectors	Eq. (7.57)
ϵ	(1) Algorithm efficiency	Eq. (4.39)
	(2) Energy deposited in a sensitive volume	p. 198
ϵ_i	Energy deposited in a collision	p. 205
ϵ_1, ϵ_2	Energy deposited in a sensitive volume by tracks of particle 1 and 2	Eq. (6.48)
$\epsilon_0^{(1)}, \epsilon_0^{(2)}$	Energy deposited in a sensitive volume in a collision resulting in one or two particles emerging from it	Eq. (6.38)

F, f

$F(x), F_\xi(x)$	Cumulative distribution function	Eq. (A.5)
$F(x)$	Collision density	Eq. (3.27)
$F(q\|x)$	Response function of a detector "measuring" distribution of deposited energy	Eq. (6.47)
$\vec{F}(\vec{v}\,)$	Effective force of multiple scattering	Eq. (5.232)
$F_1(x)$	First collision density	Eq. (3.196)
$F_1(x,t), F_2(x,t)$	Phase space distribution function for one and two particles	Eqs. (3.69), (3.70)
$F^{(n)}(x)$	Response function of a detector "measuring" moment n of distribution of deposited energy	Eq. (6.44)

$F_N(x,t)$	Phase space distribution function for an N-particle system	Eq. (3.68)
$f(x), f_{\xi}(x)$	Probability density function	Eq. (A.9)
$f(\cdot\|\cdot)$	Conditional probability density	Eq. (2.49)
$f(q\|D)$	Probability distribution of deposited energy at a given dose	Eq. (6.1)
$f(q\|x)$	Probability distribution of deposited energy for a particle that originated at point x of phase space	Eq. (6.34)
$f_1(q)$	Single event probability distribution of deposited energy	Eq. (6.2)
$f_S(q), f_{\tilde{V}}(q), f_{\Gamma}(q)$	Three components of distribution of deposited energy in method FD-2	Eqs. (6.70), (6.73), (6.76)
$\vec{f}_i$	Total force exerted on ith particle	Eq. (3.68)
$\vec{f}_{i,\mathrm{ex}}$	External force exerted on ith particle	Eq. (3.73)
$\vec{f}_{ij}$	Force exerted on particle i by particle j	Eq. (3.73)

G, g, Γ, γ

G	Total number of energy groups	p. 226
$G(\cdot\|\cdot)$	Green's function	Eq. (4.53)
$G^{(n)}(E), G_V^{(n)}(E), G_{\tilde{V}}^{(n)}(E)$	Detector response functions in method FD-1	Eq. (6.57)
$G_S^{(n)}(E), G_{\mathrm{in}}^{(n)}(E), G_{\mathrm{out}}^{(n)}(E)$	Detector response functions in method FD-2	Eqs. (6.71), (6.77)
g	Energy group number	p. 226
Γ	Surface	Eq. (3.37)

γ	(1) Random number uniformly distributed in (0,1)	p. 15
	(2) Mean square stopping power	Eq. (5.4)
	(3) Euler constant	Eq. (5.23)
	(4) Lorentz factor	Eq. (5.233)

H, h, η

$H_{k'k}^{(s)}$	Integral of a product of basis functions $h_{k'}$ and h_k over face s of a tetrahedron	Eq. (7.46)
$H_{k'k}^{(V)}$	Integral of a product of basis functions $h_{k'}$ and h_k over the volume of a tetrahedron	Eq. (7.40)
h_i	Height of a histogram bar	Eq. (2.81)
$h_k(\vec{r}\,)$	Basis function	Eq. (7.30)
η	Random variable	p. 16

I, i

$\bar{i}$	Mean charge	Eq. (5.212)

J, j

J	(1) Detector reading, observable	Eq. (3.28)
	(2) Jacobian	Eq. (A.16)
J_0, J_1, J_k	Bessel functions	Eq. (5.81)
j	Current	Eq. (4.32)

K, k

K	Boundary of the computational domain	Eq. (6.41)
$\hat{K}$	Integral operator	Eq. (3.169)
$\hat{K}_s$	Scattering operator	Eq. (3.43)
$\hat{K}_s^+$	Adjoint scattering operator	Eq. (3.47)
k	Parameter of Vavilov's distribution	Eq. (5.48)
$k(\cdot \to \cdot)$	Kernel of an integral operator	Eq. (3.168)

L, $\mathcal{L}$, l, λ

$L(\cdot)$	Likelihood function	Eq. (A.59)
L, U	Lower and upper boundaries of a confidence interval	Eqs. (A.70), (A.71)
$\hat{L}$	Boltzmann operator	Eq. (3.42)
$\hat{L}^+$	Adjoint Boltzmann operator	Eq. (3.47)
$\mathcal{L}$	Laplace transform	Eq. (6.18)
l	(1) Distance, free path	Eq. (3.6)
	(2) angular momentum	Eq. (5.216)
$\overline{l}$	Mean chord length	Eq. (6.5)
λ	Parameter of Landau's distribution	Eq. (5.29)
λ_1	Parameter of Vavilov's distribution	Eq. (5.47)
λ_i	Likelihood ratio	Eq. (2.89)

M, m, μ

m_e	Electron mass	p. 29
μ	(1) $\cos\theta$	Eq. (3.105)
	(2) mean value	Eq. (A.40)

N, n, ν, ∇

N	Sample size, number of trials	p. 259
$N(\mu, \sigma^2)$	Normal distribution	Eq. (A.40)
N_V	Number of points in V	p. 6
$n(x)$	Number density	Eq. (3.18)
n^*	Total number of bound states	Eq. (5.217)
n_e	Electron density	Eq. (5.218)
n_{bl}	Number of bound state with angular momentum l	Eq. (5.216)
n_μ, n_ϕ	Number of quadrature points for integrals over μ and ϕ	Eq. (7.27)
$\vec{n}, \vec{n}_\Gamma$	Normal vector	Fig. 3.3
ν	Number of states in a Markov chain	Eq. (3.176)
$\vec{\nabla}, \vec{\nabla}_r$	Differential operator, $\{\partial/\partial x, \partial/\partial y, \partial/\partial z\}$	Eq. (B.2)
$\vec{\nabla}_p$	Differential operator in momentum space $\{\partial/\partial p_x, \partial/\partial p_y, \partial/\partial p_z\}$	Eq. (3.68)

P, p, π, Φ, ϕ

$\mathrm{P}\{\cdot\}$	Probability of an event	Eq. (A.1)
$\mathrm{P}\{\cdot\|\cdot\}$	Conditional probability	Eq. (A.2)
P_n	Legendre polynomial	Eq. (3.145)
P_l^m	Associated Legendre polynomial	Eq. (3.149)

p	(1) Probability	Eq. (1.1)
	(2) Magnitude of momentum	Eq. (7.56)
$\vec{p}$	Momentum	p. 49
$p(\cdot \to \cdot)$	Transition probability of a Markov chain	p. 264
$p_s(\cdot)$	Probability of survival for a Markov chain	Eq. (A.79)
$\pi(x)$	Distribution of initial state of a Markov chain	p. 264
Φ	Fluence	Eq. (3.16)
Φ^+	Adjoint function, importance function	Eq. (3.33)
Φ_0	Unscattered fluence	Eq. (3.86)
Φ_g	Fluence integrated over energy group g	Eq. (7.4)
Φ_i	Charge state distribution of fluence	p. 183
Φ_n	Legendre coefficient for fluence	Eq. (3.146)
Φ_Γ	Fluence on surface Γ	Eq. (3.90)
$\Phi_{l,m}$	Spherical harmonic coefficient for fluence	Eq. (5.129)
Φ_{gl}^m	Spherical harmonic coefficient for fluence in energy groups formalism	Eq. (7.12)
ϕ	Azimuthal angle	Fig. 2.9
$\phi(\lambda)$	Landau's function	Eq. (5.31)

Q, q

Q	Total energy deposited or lost	p. 147
Q_i	Multiplicative weight	Eq. (3.175)
$\vec{\tilde{Q}}_{l,m,l',m'}$	Vector matrix in Lewis's theory	Eq. (5.131)
q	(1) Particle charge	Eq. (3.78)
	(2) fractional charge	Eq. (5.211)

R, r, ρ

R	(1) Radius	p. 37
	(2) range	Eq. (3.120)
$R(x)$	Residual	Eq. (7.34)
R_B	Radius of a charged particle trajectory in magnetic field	Eq. (5.231)
$\hat{R}_{[\vec{\Omega}',\vec{\Omega}_0]}$	Rotation operator	Eq. (4.57)
r	(1) Radial coordinate in three dimensions	Eq. (2.67)
	(2) Number of degrees of freedom of χ^2 distribution	Eq. (A.49)
$\vec{r}$	Radius vector, particle position	p. 49
$\vec{r}_0$	Initial position	Eq. (5.229)
r_s	Radius of a sphere that contains one electron on average	Eq. (5.218)
ρ	(1) Radial coordinate in two dimensions	Eq. (2.64)
	(2) Mass density	Eq. (3.31)

S, s, σ

S	(1) Source function	Eq. (3.2)
	(2) Surface area	Eq. (6.6)
$\widetilde{S}$	Biased source function	Eq. (3.195)
$S_{e\gamma}$	Collisional source of photons produced in electron interactions	Eq. (7.52)
$S_{\gamma e}$	Collisional source of electrons produced in photon interactions	Eq. (7.53)
S_g	Source function integrated over energy group g	Eq. (7.5)
$\mathrm{Si}(y)$	Sine integral	Eq. (5.57)
s^2	Sample variance	Eq. (A.58)
σ	(1) Total cross section	Eq. (3.3)
	(2) Standard deviation	Eq. (A.37)
$\widetilde{\sigma}$	Biased cross-section	Eq. (3.207)
$\sigma_<$	Cross section for soft collisions	Eq. (3.129)
σ_1, σ_2	Cross section for an interaction resulting in 1 or 2 particles emerging from it	Eq. (3.62)
σ_{ED}	Energy deposition cross section	Eq. (3.31)
σ_g	Cross section in energy groups formalism	Eq. (7.6)
σ_i	Ionization cross section	Eq. (3.29)
$\sigma_{i'i}$	Cross section for transition from charge state i' to charge state i	Eq. (5.214)
$\widetilde{\sigma}_k$	Parameter of Goudsmit-Saunderson distribution	Eq. (5.121)
$\sigma_{\max}$	Maximum cross section	Eq. (4.12)
σ_n	Legendre coefficient for scattering cross section	Eq. (3.147)
σ_s	Scattering cross section	Eq. (3.12)
$\sigma_s(g' \to g)$	Differential scattering cross section in energy groups formalism	Eq. (7.7)
$\sigma_{s,e\gamma}$	Scattering cross section for electron interactions producing photons	Eq. (7.52)
$\sigma_{s,\gamma e}$	Scattering cross section for photon interactions producing electrons	Eq. (7.53)
$\sigma_{sl}(g' \to g)$	Legendre coefficient for differential scattering cross section in energy groups formalism	Eq. (7.15)
σ_{tr}	Transport cross section	Eq. (3.14)

T, t, τ, Θ, θ

t	(1) Distance, step size	p. 66
	(2) time	Eq. (5.222)
$\tau(\vec{r}_1, \vec{r}_2)$	Optical distance between points $\vec{r}_1$ and $\vec{r}_2$	Eq. (3.10)

$\hat{\Theta}_a$	Estimator of parameter a	p. 260
θ	Polar angle	Fig. 2.9
$\langle\theta_0^2\rangle$	Mean square scattering angle	Eq. (3.138)

V, v

V	Volume	p. 6
$\bar{V}$	Complement volume	Eq. (3.53)
$\tilde{V}$	Extension volume	Eq. (6.54)
V_B	Bounding volume	p. 6
$V(r)$	Interaction potential	Eq. (5.216)
$\mathrm{Var}\{\cdot\}$	Variance	Eq. (A.33)
v	Velocity	Eq. (5.221)
v_0	Bohr velocity	p. 185
$\vec{v}_0$	Initial velocity	Eq. (5.226)
v_1	Ion velocity	Eq. (5.219)
v_e	Velocity of orbital electron	Eq. (5.219)
v_F	Fermi velocity	Eq. (5.220)
$\bar{v}_r$	Average relative velocity	Eq. (5.219)

W, w, Ω, ω

w	(1) Energy of ejected electron	Fig. 2.8
	(2) weight	Eq. (3.195)
$w_j^{(C)}$	Weight in Chebyshev quadrature	Eq. (7.24)
$w_i^{(L)}$	Weight in Legendre quadrature	Eq. (7.23)
ω	Solid angle	Eq. (2.70)
ω_B	Gyration/precession frequency	Eq. (5.224)
$\vec{\Omega}$	Unit vector, defines a direction in three dimensions	p. 49

X, x, ξ, χ

$x_{1/2}$	Median	Eq. (A.53)
$x_{1/4}, x_{3/4}$	Lower and upper quartiles	Eq. (A.54)
x_∞	Equilibrium depth for charge exchange processes	Eq. (5.215)
x_q	Quantile	Eq. (A.55)
ξ	Random variable	p. 16
$\xi_{1/2}$	Sample median	Eq. (A.57)
$\bar{\xi}$	Sample average	Eq. (A.56)
χ^2	Chi squared statistic	Eq. (A.48)
χ_a, χ_c	Parameters of Molière distribution	Eqs. (5.105), (5.106)

Y, y

$Y_{l,m}(\vec{\Omega})$, $Y^*_{l,m}(\vec{\Omega})$	Spherical harmonic and its complex conjugate	Eqs. (5.128), (5.129)
y	Lineal energy	Eq. (6.5)
$\bar{y}_F$	Frequency mean lineal energy	Eq. (6.8)
$\bar{y}_D$	Dose mean lineal energy	Eq. (6.13)

Z, z, ζ

Z_1	Atomic number of an ion	Eq. (5.211)
z	Specific energy	Eq. (6.4)
$\bar{z}_F$	Frequency mean specific energy	Eq. (6.9)
$\bar{z}_D$	Dose mean specific energy	Eq. (6.14)
ζ	Random variable	p. 16

Index

O.N. Vassiliev, *Monte Carlo Methods for Radiation Transport*, Biological and Medical Physics, Biomedical Engineering, DOI 10.1007/978-3-319-44141-2